

Nuclear powers must act quickly

FOR the next few months the question of ratification of the SALT 2 treaty by the US Senate is bound to occupy the attention of most Western arms controllers and politicians with an interest in arms control. Then early next year, regardless of the outcome of the SALT debate, their thoughts will have to be diverted to the five-yearly review conference of the Non-Proliferation Treaty (NPT) planned to take place in May 1980. And it is certain that the NPT conference will be the occasion for yet more bitter questioning by the non-nuclear-weapon states of the seriousness of commitment of their nuclear counterparts to Article VI of the NPT, which said that all parties to the treaty undertook to "pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament . . .".

A successful outcome to the SALT debate would certainly tone down the criticism to a certain extent. But it is clear that what the non-nuclear-weapon states really want is a weapons-test ban. Diplomatic wording of the agreed statement at the end of the 1975 NPT review conference could not conceal exasperation at continued testing. "A considerable number of delegations" wanted a ban for a specified time, with a view to negotiating eventually a "universal and permanent cessation of all nuclear weapons tests". Such a treaty has yet to come about.

True there is now US-Soviet agreement to restrict tests to a maximum of 150 kilotons yield. (The UK uses the US test-site for its one or two explosions per year so is also bound by this agreement; the French and Chinese are not.) True there are US-UK-Soviet talks proceeding at present on a possible short-term treaty. True there has been some softening of the previous Soviet intransigent position on monitoring stations on its own soil. But there is as yet little evidence that there will be a solid treaty by the time of the NPT conference. Two points are as yet unresolved.

The first is rather silly. Some months back the Soviet Union proposed that for the three-year period, now regarded as the most that a ban would be negotiated for, each party to the tripartite talks would have ten seismic monitoring stations on its soil to verify compliance — the data to be mutually available. This was a big step forward for the Soviet Union, but did they really mean ten stations on British territory? Apparently they did. British territory does not amount to much these days outside the UK. There would be no scientific point in more than two stations in the UK to monitor compliance, and for the rest the instruments would have to be placed in Hong Kong (useful

for keeping an eye on the Chinese, were the noise levels not so high) and sundry remote islands, probably equally noisy and therefore useless. As the monitoring posts are supposed to be the best that technology allows, the bill for this futile operation could not be less than tens of millions of pounds. The British, who on the whole wanted to take a back seat in these negotiations and not get in the way of a superpower agreement, suddenly find themselves pushed up front in rather a ridiculous way. Scientific honesty requires that they resist this pointless waste of resources; political necessity requires that they not let the Soviet Union make capital out of it.

The second sticking point is more substantial. After a three-year total ban, what comes next? On the surface the question is what to call the review conference at the end of the period, but the real issue is whether a three-year ban could ever be converted into a longer one. Yes, say the Russians, if the Chinese and French behave themselves and if some way can be found of coping with 'peaceful' explosions (a thorny problem). No, say the Americans, we can't be tied; there is enough domestic opposition as it is to any ban at all — the weapons laboratories are pushing the issue of stock-pile testing harder year by year and claiming (not without dissenting voices) that weapons untested for long periods will be unreliable. Aldermaston, with its trans-Atlantic ties, would say the same if allowed to be asked, but the British are burying any internal differences in a united front of keeping the options open for longer term ban. Mrs Thatcher is known, however, to be showing some interest in the UK test ban position; it is on the cards that Britain might eventually show less enthusiasm for long term restraint. Whatever agreement is finally reached, it is indisputable that in the US, the President will have to sell it very hard by means of vigorous 'safeguards' such as more, not less, financial support for the weapons laboratories.

Time, it hardly needs saying, is running out for a convincing treaty to be prepared before the NPT conference, especially as the multilateral Geneva Disarmament Conference will want to take a look at it. Non-nuclear-weapon states are, rightly, going to work up a head of steam over the failures to prevent vertical proliferation whilst they have been restrained from horizontal proliferation. Their position of restraint is not made easier by the continual gossip about the nuclear capabilities of South Africa, Pakistan and Israel. Unless the three negotiating nations put a lot more goodwill into their discussions in the next few months they will deserve all the obloquy they will certainly get. □



British scientists wait for crucial decision that could ease pressure on promotions

In two weeks the Science Research Council is to make a decision on a controversial plan aimed at easing the critical staff stagnation problem which now affects universities in the UK. On 18 July, its central council will decide whether or not to give the go-ahead to a special release fellowship scheme that is intended to help more senior research workers vacate tenured posts which can then be filled by young scientists.

The age structure at UK universities has been deteriorating since the higher education expansion of the early 1960s, when new departments were created and were manned at senior levels with young scientists. In recent years, economic recession has ended the creation of additional posts and the distribution of posts has been frozen.

Indeed, in some science departments there have been no new appointments for the past ten years. Where there have been vacancies these have been filled from within departments, thus halting the crucial flow of ideas and talent between universities.

Now, in an attempt to stop this stifling of fresh talent, the SRC is seriously considering the establishment of special fellowships which would fund research workers advanced in their careers and allow them to leave university posts to concentrate solely on research without administrative responsibilities. Their vacated positions would then be filled by young scientists and over subsequent years allow the phasing-in of new research workers.

However it is stressed that the scheme is not meant as a form of premature retirement for ageing researchers, and only those who are active and still have important scientific contributions to make will be considered for funding.

The proposal to be discussed by the SRC this month is for 15 fellowships to be set up in the first year of the programme with further annual increments of 10 new awards. This could then provide a total of 45 special release fellowships within four years.

But there are several difficulties associated with this plan. For one thing, it

has not yet been agreed whether the SRC should pay for the senior researcher's salary or if it should pay for the junior post. The difference could represent a variation of about 50% in final costs and consequently estimates vary between £300,000 and £450,000 a year for the ultimate price of the full fellowship scheme.

However the council is at present particularly concerned about its future budget allocations and may be very reluctant to commit itself to such a level of funding. At its meeting, it may decide to cut the numbers of fellowships or may even postpone any decision until September when it hopes to learn its financial guidelines for the coming years.

Indeed, some council members are very cautious about the whole idea of the scheme, believing it to be the role of the University Grants Committee to provide proper staff structures within departments. Combined with these other factors, this view puts the council's coming decision very much in the balance.

Robin McKie

Dearest oil puts brake on Third World growth

THE effect of the current energy crisis on Western nations has attracted considerable international attention but precious little has been said about the non-OPEC Third World countries. Energy experts who gathered at a forum on 'Third World Energy Strategies and the Role of Industrialised Nations' organised recently (20-22 June) by the Royal Institution in London, described an extremely critical energy situation developing shortly in these countries, which could almost threaten their very survival.

Their predicament was well illustrated by a case study of Kenya presented by T. S. Tuschak, a UN energy consultant based in Nairobi. Mr Tuschak pointed out that Kenya consumed in 1978 a little less than 12 million barrels of oil (equivalent to about one week's consumption in Canada). This small quantity of oil accounted for about 85% of the energy consumed by the country's modern sector. About 80-82% of this oil was consumed in essential sectors like industry, agriculture and goods transport. The rest was accounted for by private automobiles, which too in many cases were not inessential because of an underdeveloped public transport.

The possibility of oil savings in the essential sectors is at best 7%. Consumption by private transport can at best be reduced by about 10%. Kenya, therefore, has few options but to maintain oil imports at near current levels, and will possibly even have to increase them if the economy continues to grow.

Tuschak estimates that for \$1/barrel increase in oil price, the country's export earnings would have to increase by about 1% over the 1978 level. This expansion in exports may sound small but could well be impossible considering the global slow down in economic growth.

Coffee and tea represent 50-60% of Kenya's total exports and they are about the only commodities in the country capable of any significant expansion in world markets. Their export would have to increase by as much as 30% in a year to cover the incremental foreign exchange needs. "The only way to meet this bill," said Tuschak, quite despondently, "is to further impoverish the country, by paying for oil by funds destined for development". Few developing countries will be able to afford an oil price of \$20-22/barrel.

The other possibility, of course, is for Western countries to bale out the developing countries. But this option brought forth only a spirited discussion, virtually no agreement.

Stephen Bosworth, the US Ambassador to Tunisia, and former chairman of North-South negotiations on energy at Paris, plainly warned Third World energy planners that Western aid and loans from international financing institutions are unlikely to rise fast enough to meet the needs generated by oil price rises. Nor is an oil sharing system which gives preferences to developing countries is likely to come into operation he said. The world will have

a slower rate of growth in the future. "We have seen that Western parliaments react unfavourably to any real transfer of resources in those conditions."

The discussions on alternate energy sources that could replace oil revealed two important Third World concerns: first, that the Third World is extremely unhappy with the constant preaching by Western scientists and technologists that unconventional energy sources like wind and solar power, and fuelwood plantations are best for it. All Third World energy experts gathered in London agreed that these energy sources would be good for developing countries but they would first want them to be fully tried and tested.

Secondly, they pointed out, none of these sources can provide an alternative to oil in the short-term which is really the critical period for the non-oil producing Third World.

Dr A. K. N. Reddy of the Indian Institute of Science at Bangalore threw an outright challenge to the developed countries: "Teach us to live without oil but by doing it yourselves". That the developed countries must first put their own house in order was a demand that was echoed by many Third World participants.

Dr Joao A. Meyer, of the State University of Campinas in Sao Paulo, criticised Western efforts for following their traditional approach to technology transfer even in the new field of non-traditional energies. They are once again trying to sell gadgets, instead of trying to



Fledgling transport systems could be hit by petrol shortages

help developing countries solve their own problems, he said; the only difference now is that the accent is more and more on solar power plants instead of nuclear plants.

"Often we do not even know whether immature energy technologies are being pushed upon us", Dr Meyer alleged. "And where we have developed technologies indigenously, for instance, the production of alcohol from sugarcane, we find ourselves besieged with claims of patent infringements and demands for royalties."

Among alternate energy sources, discussions centred on wood, which provides over three-quarters of the Third World's population with its energy needs. For these people, the wood crisis is worse than the oil crisis, and certain to become even more acute. Kenyan experts revealed that wood scarcity in their country is already so acute that serious thought is being given to providing electricity for cooking in rural areas to replace the use of firewood. "Until 1973, we thought we had the wood problem licked", said India's energy planner Sankar. "We were subsidising the use of kerosene in villages to replace the use of firewood. But then in 1973 the carpet was pulled out right under our feet."

The answer is to grow more wood; but the forum discovered that first, scientists themselves do not have adequate knowledge about fast growing trees and forestry techniques, and second, even if existing knowledge could be applied right away, a British expert pointed out, it would be 1996 before rural areas could get wood. Problems of social organisation in Third World villages will prevent rapid diffusion of innovations. India's progress with biogas plants to meet rural cooking energy needs has also been very slow.

There was, however, total unanimity on the need for developing countries to plan

their future energy use carefully and to discover new sources. Once developing countries identify their energy problems, and choose the technologies they want to solve them, then the developed countries can help them in developing those technologies instead of passing on to them packages that imitate successful ones in the West, said Professor Ignacy Sachs of the University of Paris.

However, the forum learnt that though very soon most Third World countries would be setting up ministries of energy, few of them had either the relevant information systems or the trained personnel with analytical capabilities to make proper economic, social and technical choices in the energy sector. It is here that developed countries and international organisations can help most, felt the Third World participants.

Professor M. S. Wionczek of Mexico suggested the establishment of an institute for advanced studies in energy policy, possibly similar to the Institute of Theoretical Physics at Trieste, a unique institution which provides scientists from developing countries an opportunity to meet and discuss with leading scientists, including Nobel laureates from developed countries. Venezuela's Minister of State for International Economic Affairs, Mr M. Perez-Guerrero, suggested a new institute be set up within the UN to help Third World countries in energy planning.

But there were no solutions to the problem that Third World experts had themselves identified: what do developing countries do to solve their energy problems in the immediate future? As India's Sankar put it: "I find no dearth of experts prepared to chart out an energy strategy for the year 2000. My problem, however, is 1980 and 1981. Who will help me solve that?"

Anil Agarwal

Experts agree on Mediterranean pollution treaty

Scientists from 14 of the 18 Mediterranean states unexpectedly agreed last week on the text of a treaty to control land-based sources of pollution — but the rate of implementation of the treaty will depend on how fast scientific work can be completed.

"The deadlock is over" said the director of the United Nations Environment Programme's regional seas unit, Stjepan Keckes, this week, "but it is not the breakthrough that some newspapers have reported".

One of the main obstacles had been a disagreement between the northern, industrial nations, and the southern underdeveloped states, over whether to impose uniform emission standards on polluters, or to define common goals for coastal water quality. The underdeveloped states argued that the former would penalise the growth of new industry in an area which presently had little industry and little pollution; while the north contended that the latter would allow the south to develop cheap and heavily polluting industries which would compete with more expensively controlled northern firms. In the end the underdeveloped states won the argument, and the treaty specifies coastal water quality standards. "This is the most logical arrangement" Keckes told *Nature*. "It is the same pragmatic scheme you operate in Britain".

The treaty will go to a specialist diplomatic meeting in Athens in April for signature. Ratification comes later.

"These are the first steps", said Keckes, a marine scientist who is also in charge of the scientific assessment of the state of the Mediterranean. "But we don't have enough evidence yet. We can define water quality for bathing, using a microbiological standard — we've been measuring on 150-200 beaches — and later we hope to define heavy metal limits for shellfish growing" he said. "But the rate of implementation of the treaty will depend on how fast we can define standards scientifically."

Progress is therefore limited by the money UNEP can spend on its pollution monitoring programme, 'Medpol'. Earlier this year, a full meeting of the contracting states voted to delete Medpol's air pollution research programme (see 15 February p 506) — whereas air, Keckes feels, may be the main source of the central Mediterranean's metallic pollution.

"The research programme needs an additional boost and additional money" said Keckes. He hopes to present a strong case for restarting the air pollution programme to the next full meeting of contracting states in February 1981.

Robert Walgate

Proxmire must answer libel charge

THE US Supreme Court last week decided that Senator William Proxmire could not claim Congressional privilege in making derogatory remarks about a scientist's research in press releases and news broadcasts announcing his Golden Fleece award. The award is presented regularly to esoteric-sounding research projects supported by federal funds.

The court also decided that, in claiming libel damages for remarks made by the Senator in presenting such an award, a scientist should not be considered a "public figure" by virtue of having received public funds, and therefore was not required to demonstrate that the remarks had been made with deliberate malice.

The Supreme Court's decision was based on an appeal made by Dr Ronald R. Hutchinson of Kalamazoo, Michigan, in connection with a libel suit that he had filed against the Senator following the award of a Golden Fleece to the National Aeronautics and Space Administration for supporting Dr Hutchinson's research on the latent aggression of laboratory monkeys.

Labelling the research as being to study "why monkeys clench their teeth", Mr Proxmire claimed in a speech to the US Senate in 1975 that "the good doctor has made a fortune from his monkeys and in the process made a monkey out of the American taxpayer", a charge he repeated in a newsletter to his constituents and a subsequent radio television interview.

Dr Hutchinson subsequently filed a libel suit, claiming that Mr Proxmire's actions had caused him to be held up to public scorn and resulted in a loss of respect in his profession, as well as "loss of income and ability to earn income in the future".

The suit was rejected by a lower court

partly on the grounds that Dr Hutchinson had become a "public figure" by soliciting government grants and publishing articles in professional journals, which required him to demonstrate "actual malice", and partly because Senator Proxmire claimed legislative privilege in making the remarks.

In last week's decision, however, the Supreme Court stated that, although local newspapers had reported Dr Hutchinson's receipt of research grants, and although he was quoted in the press in reacting to the Golden Fleece award, neither of these was sufficient to make him a "public figure".

"His access [to the media] such as it was, came after the alleged libel and was limited to responding to announcement of the award. Those charged with alleged defamation cannot, by their own conduct, create their own defence by making the claimant a public figure," the court announced.

On the question of legislative privilege, the court ruled that this could not be used by the Senator as a defence, since there was no evidence of past Congressional concern to grant absolute privilege for remarks made outside the legislative chambers, since neither the news-letters nor the press release were "essential to the deliberation of the Senate", and neither was protected as part of the responsibility of members of Congress to inform the public of their activities.

In the light of these decisions, the Supreme Court directed that the case be returned to the Court of Appeals "for further proceedings consistent with this opinion". In a dissenting opinion, Mr Justice Brennan declared that in his opinion "public criticism by legislators of unnecessary governmental spending whatever its form, is a legislative act shielded by the Speech or Debate clause". □

NASA bets 50-50 on space shuttle next summer

OFFICIALS of the US National Aeronautics and Space Administration told a Congressional committee last week that the first orbital flight of the reusable space shuttle is not now likely to take place until next summer at the earliest, and that this may mean at least four satellites scheduled for early shuttle launches will have to seek alternative ways of getting into orbit.

Dr Robert Frosch, the agency's administrator, told the House Science and Technology Committee that technical delays in the development of the shuttle's engines and in fitting its heat-resistant surface tiles meant that there was now only a one-in-five chance of launching the shuttle by the end of the first quarter of 1980, a launch originally planned to have taken place this spring, and a 50-50 chance of a first flight in the second quarter.

"This slip means our first operational flight, which had been planned for

February 1981, will be delayed until September 1981," Dr Frosch said. "This means some of our early users will have to shift to use the backup expendable launch vehicles to get their satellites into orbit." (It also means that it will now be necessary to revise the launch schedule for the European Space Agency's Spacelab, which was due to be launched by Space Shuttle in August 1981).

Among the satellites whose launches will be affected by the new shuttle delays is a geosynchronous operational environment satellite which was planned to be launched by the National Oceanographic and Atmospheric Administration; shifting from a shuttle launch to an expendable launcher could increase the cost from about \$10 million to \$22 million, and the same would apply to two commercial communications satellites that were to have been among the first payloads. □

New warning on risk from antibiotics in animal feeds

THE US Congress' Office of Technology Assessment reported last week that antibiotic resistance is continuing to increase in the US population, and warned that, although quantitative relationship remains uncertain, the continued use of antibiotics in animal feeds seems to be contributing to this process.

The OTA's warning came in a report on "Drugs in Livestock Feed" prepared at the request of the Congress following the Food and Drug Administration's proposal in 1977 to ban the use of penicillin and sharply restrict the use of two tetracycline drugs in animal feed. According to the industry, sales of antibiotics and similar antibacterial agents added to animal feed exceeded \$180 million last year; and the OTA estimates that the FDA's proposed restrictions could increase retail meat and poultry prices by between 3% and 28% in one year.

However the OTA also says that antibiotics in animal feed contribute to a "growing pool of drug-resistant bacteria", and that physicians are now reporting reduced effectiveness of the drugs in treating disease; and it points out that many of the drugs on which restrictions are proposed could be replaced with alternative drugs that are already approved by the FDA.

It is up to Congress, the agency concludes, to weigh the economic costs of banning the drugs in animal feeds against the health risks of allowing their continued use. "These decisions involve value judgements that cannot be based simply on monetary considerations. And the lack of scientific certainty on the magnitude of both the probable health risks and the attributed increases in meat production makes the formulation of a balance-sheet approach difficult," the OTA report says.

The short-term consequences of a ban on the drugs could be significant, since production might decrease, leading to higher prices. But the OTA says that the long-term consequences are less certain, probably resulting in small decreases or no changes in production and small increase in both consumer prices and overall producer incomes.

Dr Joyce Lashoff, OTA's assistant director for health, emphasised at a press conference last week that the OTA made no specific recommendations about how Congress should act, merely listing a range of options. Giving a personal view, however, she said that she believed the availability of substitutes meant that the economic impact of the regulations proposed by the FDA would not be great enough to offset the potential risk to human health if the use of antibiotics remained unrestricted. □

news in brief

New agency proposed to handle nuclear waste: A loss of confidence in the US Department of Energy's ability to come up with a co-ordinated nuclear waste disposal programme means that the problem should be handed over to a new committee of federal and state representatives, the General Accounting Office, the investigatory arm of the US Congress, announced last week. In a report prepared at the request of Senator John Glenn, the GAO pointed out that a number of states had already passed various forms of legislation concerning how nuclear waste issues should be handled, but that the federal government had so far failed to come up with an acceptable permanent storage plan for high level wastes. This was becoming a growing obstacle to the development of nuclear power. "We believe it is very unlikely that making the Department of Energy the responsible lead agency to plan and co-ordinate the programme will establish public trust and confidence", the GAO report states.

German mathematician fights 'Berufsverbote': The case of Horst Eckart Gross, 36, who was excluded from a job as assistant professor at the University of Oldenburg, has entered a new phase. West German authorities are now arguing that the recent public campaign in support of Gross' employment is further evidence of his "unconstitutional behaviour". Gross was nominated for a position teaching the history and professional practice of mathematics by the normal university committees in February 1977. The Verfassungsschutz, the West German body responsible for vetting political opinions of government employees, delayed the appointment for almost a year and then said in its final report that there was information which gave reason to doubt Gross' loyalty to the "Freiheitliche Demokratische Grundordnung" (free and democratic constitution). The information in question consisted of Gross' activities in a German-Cuban friendship society which has local groups in more than 30 West German cities. Members of the organisation include trade unionists and members of parliament. Gross is also accused of membership of the German Communist Party and of past membership (1967/68) in the SDS.

The council of the University of Oldenburg passed a strong resolution in support of Gross on 11 January 1978 protesting at his examination by the Verfassungsschutz and calling for his immediate employment. "This Berufsverbote case shows perhaps more clearly than any other the unconstitutional methods used by the political police", it said.

Cuba to build nuclear reactor: Cuba is proceeding with a plan to build the first nuclear power plant in the Caribbean. Mr Luis Beltran, minister in charge of the electrical power industry, announced last week that construction on the Soviet-designed reactor would start this year at Cienfuegos on Cuba's southern coast. The reactor is of the pressurised water type similar to the one involved in the Harrisburg accident. Mr Jorge Perez-Lopez, a US government economist, claims that the Soviet design lacks two safety features which are common to PWRs built in Western Europe and the US. Writing in the magazine *Cuban Studies*, Perez-Lopez says that Soviet PWRs lack a backup cooling system and an outer casing to contain possible leaks. Perez-Lopez suggests that Cuba and the Soviet Union have a mutual interest in developing Cuban nuclear power at this time. The Soviet Union conserves its oil supplies for trade on the world market and Cuba lessens its financial dependence on Soviet oil.

'Plenty of jobs in the UK': According to Neil Macfarlane, minister of science in the new Conservative government, jobs are going begging in the UK. Speaking at an Association of British Science Writers lunch, Macfarlane stated that "for people with the right skills there are plenty of jobs". When asked whether this included scientists on short-term contracts, Macfarlane declined to comment, agreeing with his host, Roger Lewin, deputy editor

of *New Scientist*, that "the issue was too complicated for a brief discussion".

Meanwhile, the Department of Education and Science has just released a recruiting pamphlet, *Science at Work*, intended to attract school students to science careers. The pamphlet is one of a DES series whose "main emphasis is on educational courses and qualifications that lead to jobs". It will be circulated to all schools, colleges and careers offices in the UK. The booklet specifically cites teaching as a career option. "To ensure the continuity of science progress, there is always a need for scientists as teachers in schools, colleges, polytechnics and universities." Also listed are possible careers in microelectronics and librarianship. *Science at Work*, Department of Education and Science, Room 2/11 Elizabeth House, York Rd, London SE 7PH. Free.

Europeans are afraid of new science and technology:

A public opinion poll shows that 80% of the people in the EEC fear for the future of the world they live in because of the "despoiling of natural life and the countryside through pollution of all kinds". A second major fear (67%) is of widespread unemployment caused by automation. A majority (53%) also expressed concern about new technologies in housing, traffic and food and the introduction of drugs



"And just how long have you had this fear of unemployment caused by automation?"

that "may severely affect the human personality". The study had been requested by the EEC Commission from European Omnibus Survey, a group of eight specialised institutes in the member countries. Four countries — Germany, Denmark, the Netherlands and the UK — rank pollution as their number one concern and four others — France, Belgium, Ireland and Luxembourg — rank unemployment first. The Italian public fears an overall deterioration of living conditions. This includes the conviction that science is not being used for the benefit of the population as a whole.

The poll shows that even though only 10% considered themselves sufficiently educated to discuss science competently, the majority insisted on being involved in major development choices. The nuclear power questions show that the public has been convinced that curtailing nuclear power could mean loss of electricity. But at the same time, 36% are still opposed to nuclear power compared to 44% who favour it. The poll disagrees sharply with a 1977 survey, which showed "clearly" that there was no mistrust of science. *The European Public's Attitudes to Scientific and Technical Development*. Commission of the European Communities. Brussels.

ETSU finds most renewable energy sources safe: The UK's Energy Technology Support Unit has given an environmental clean bill of health to most renewable energy sources. In a preliminary report to the Department of Energy, the unit finds that the impact of solar and wind energy conversion is "likely to be relatively small". The impact of tidal energy and wave energy is considered to be more complex. Wave energy converters could interfere with herring and salmon migration paths and with ship navigation.

The report lists 13 areas of concern about tidal energy conversion, including effects on ports, freshwater fisheries, pollution dispersion and agriculture; £1.5m has been devoted to the further study of these effects. *Environmental impact of the renewable energy sources*. ETSU, Harwell, OX11 0RA.

Alpha-fetoprotein screening: too hot to handle?

US manufacturers want to market pre-natal screening kits for anencephaly and spina bifida. But can US medicine cope with the risks? **David Dickson** reports

WHEN Dr Donald Kennedy returned to Stanford University last Friday after a two year spell as Commissioner of the Food and Drug Administration, he left unresolved a classic but little-publicised example of the problems society faces with potentially hazardous technologies: namely whether — and if so under what circumstances — the FDA should approve the marketing of techniques capable of warning an expectant mother that she is bearing a child with a serious birth defect.

The techniques in question use a radioimmunoassay to measure the level of alpha-fetoprotein (AFP) in the mother's blood serum. A high level of AFP indicates that the foetus has an above-average — although still relatively small — chance of suffering from a neural tube defect (NTD), manifest at birth as either anencephaly or spina bifida. A precise diagnosis can subsequently be achieved using ultrasonography and amniocentesis.

If properly used, this series of tests could theoretically pinpoint the one or two children in a thousand born in the US with a neural tube defect at a sufficiently early stage of pregnancy to permit an abortion, if the mother so chooses. A wrong diagnosis, however, or a misinterpretation of preliminary test results, could lead to the abortion of a healthy foetus.

It is a situation in which even obstetricians have expressed doubts about their present ability adequately to handle the whole testing procedure. The FDA, to which a number of manufacturers have submitted applications for licences to market AFP screening kits, now faces the problem of building sufficient checks into the licensing process to minimise the risks involved. In doing so, some fundamental questions have been raised about the ability of an essentially free-market medical system to provide the coordination, support and oversight needed to make this possible.

The discovery that neural tube defects are associated with high levels of AFP in the amniotic fluid was first made in the early 1970s by two scientists at the University of Edinburgh in Scotland. It is thought to occur because the defect allows the protein, which is only found in foetal blood, to leak into the surrounding amniotic fluid.

Early tests showed that measuring AFP level by amniocentesis provided a highly accurate guide to the chances of a foetus suffering an NTD. And subsequent tests in the United Kingdom also indicated that merely measuring AFP levels in the mother's blood could provide an initial clue to such a likelihood.

The problem, however, is that a correct diagnosis based initially on serum testing requires a number of separate further steps, each of which has to be carried out with a high degree of accuracy. Firstly the AFP level in the maternal serum must be measured between the 16th and the 18th week of pregnancy; if the AFP reading is abnormally high (typically considered to be in the top 5% of mothers tested), a second measurement is made, although the woman still has only a one-in-fifty chance of carrying a defective foetus.

A second high reading, indicating a 10% chance of a defect, leads to further testing to discover the reason. Ultrasonic examination of the womb can usually indicate if the foetus is anencephalous, or if the high AFP is caused by twins or premature death. If none of these is found, amniocentesis — which is not in general applied earlier, since the risk of accidentally inducing an abortion is higher than the general one-in-a-thousand risk of bearing a defective child — is used as a final check.

A \$25 million business

In theory this complete screening procedure, if made generally available, could lead to a significant decrease both in the 2000 to 3000 children who are born and survive every year in the US with spina bifida, and in the anguish to the parents of those (in particular the anencephalics) who do not survive at birth. A national programme has therefore been urged by many, and is already being studied in Britain, to minimise what is currently the most common serious birth defect. Manufacturers envision a multi-million dollar market for their screening kits and reagents; and private laboratories a comparable market for their testing services.

Three main concerns, however, have been raised by the prospect of the techniques being freely marketed and becoming generally available. The first is the general opposition of the politically powerful anti-abortion lobby, critical of what one Right to Life committee member last week described as any "search and destroy" mission involving pre-natal screening.

The second concern is on the economics of mass screening, and revolves around cost-benefit arguments. Research workers at the Center for Disease Control at Atlanta, Georgia, have calculated that a screening programme offered to all pregnant mothers in the US, with a 40% acceptance, would cost about \$25 million a year; but the averted annual cost would be about \$50 million. Others challenge these figures, arguing that the risks are insuffi-

ciently known to enable realistic quantification.

It is the third consideration, however, which is presently causing the greatest headache to FDA officials, namely how to ensure adequate quality control of the whole process, and not just of the screening kits, to minimise the risks involved. For as one FDA official points out: "The use of the test is so integral to obstetric practice that it is hard to look at this purely as a technique without also taking into account the way that it is going to be used."

Such concerns, focused in particular on the dangers of causing large numbers of unnecessary abortions, have four main aspects:

- how to ensure adequate counselling for the parent, both with respect to gaining initial "informed consent" and to correctly interpreting the test results at various stages;
- how to ensure physicians are sufficiently informed about the techniques to be able to interpret laboratory results properly, and provide follow-up guidance;
- how to ensure that the complete range of testing facilities is available to every mother taking the test, so that a decision to abort, for example, is not made solely on a positive first test due to the lack of sonography or amniocentesis facilities (or an inability to pay for both); and
- how to ensure sufficient accuracy in the information provided to physicians by testing laboratories.

Although such issues have been debated in academic circles for a number of years, alarm bells only started ringing at the FDA last Autumn after a hearing of a review panel of the agency's Bureau of Medical Devices on applications to market screening kits which have been developed by a number of manufacturers, including Hoffman la Roche, Abbott Laboratories, and the Wampole Division of Carter-Wallace Inc.

Shortly after the hearing, Mrs Carol Bucholz, a director of the Spina Bifida Association, wrote a strongly-worded letter to Dr Kennedy complaining that, as "the obvious consumer population" the association had been neither informed of nor represented at the meeting.

Members of the association mounted a vigorous letter-writing campaign to the FDA. And in their opposition to unrestricted licensing, they were soon joined by other groups and institutions. The director of the Center for Disease Control, for example, Dr William H Foege, wrote that without proper control, AFP screening could lead to unnecessary abortions and false-negative diagnoses which would "continue to be made until successful malpractice suits alter medical practice to stop these mistakes"

The Health Services Administration an-

nounced that it did not feel adequate facilities existed to provide the necessary sonography and amniocentesis services if the screening technique became suddenly available on a national basis. "We therefore recommend deferral of the approval for the production and marketing of the test kit at this time", the head of the HSA wrote to Dr Kennedy.

Perhaps most significantly, opposition to unrestricted licensing was also expressed by two professional medical bodies, the American College of Obstetricians and Gynaecologists, and the American Society for Human Genetics. ACOG admitted openly that, in addition to sharing the concerns of others, it did not feel its members were sufficiently prepared to carry the responsibility which the general availability of AFP screening techniques might impose.

"Our position is that we need to have a realistic protocol and set of guide-lines prepared both for the physician and for the patient. It is going to take institutional approaches and a lot of education to learn how to use this technique properly, and we are concerned that it should not be turned loose until we have worked out our line of approach," Dr Ervin E Nichols, director of practice activities at ACOG, told *Nature* last week.

Many of these views were aired at a further public hearing of the medical devices panel in February of this year. However the director of the Bureau of Medical Devices, Dr David Link, subsequently recommended to the Commissioner that licences be granted without broad restrictions on marketing, and subject only to requirements for labelling, reporting, and the provision of information to physicians and patients.

But things were already stirring at the top of FDA. In March the agency announced that it was rejecting an application from Abbott Laboratories to have its screening kit reclassified from a category of medical devices that require pre-market testing on an individual basis, to one which merely required compliance with pre-determined standards.

In announcing its decision, and over-riding the recommendation of an advisory panel that the reclassification be granted, the FDA stated that it was "not convinced that a standard would provide reasonable assurance of the safety and efficacy of the device" if it were reclassified as requested. It added that "the petitioner did not state how FDA could assume that ultrasonography or equivalent procedures, i.e. amniography, would be used for determining gestational age", essential information for correct interpretation of the test data.

The various groups have now met with top officials at the FDA to state their case. And it appears that, despite the continued opposition of some manufacturers, the FDA is now prepared to add restrictions to the licences that it issues. A possible

protocol is now being developed by ACOG and the American Academy of Pediatrics. And the big question is how far restrictions will go; for, given that all parties in the debate accept the need for a high degree of quality control, the issue remains of how in practice this should be ensured, in other words how sufficient public accountability can be achieved.

On several points both the manufacturers seeking licences and those seeking restrictions on such licences are broadly in agreement. For example ACOG is now working on appropriate forms of physician and technician education, with considerable support of the manufacturers (some of whom have already produced trial educational material to this end).

Who is in control?

The main source of controversy, however, is over who should control the quality control process. Manufacturers and laboratories see quality control as essentially a technical issue; they would like the federal government to agree on standards of accuracy, and then merely be required to meet these standards. Critics, however, such as the Washington-based Health Research Group, argue that this is insufficient; they want more direct public involvement and accountability, perhaps through either federal or state agencies, in monitoring the way that laboratories conduct the tests, requiring for example, regular inspections rather than occasional proficiency testing.

Direct regulation of laboratory practices, however, is taking the FDA into uncharted areas. At present, government regulation is a "loose, patch-work affair" according to Mr Robert Leflar of HRG, with responsibility for different types of laboratories divided between different agencies. The FDA is confident that it can devise an adequate control system, possibly through the CDC; but has yet to announce details of how this will be done.

One approach which, some feel, would act as a half-way house to general marketing of the screening techniques would be through carefully-controlled pilot programmes. One successful pilot programme has already been in operation for several years on Long Island, run by Dr James Macri of the State University of New York at Stonybrook; another is being planned under the auspices of the State of Maryland's Commission on Hereditary Disease, set up in 1973 to guide methods for the detection and management in hereditary disorder in Maryland.

Listing the problems that surround the accurate use of AFP screening techniques, Dr Neil A Holtzman of the Johns Hopkins Medical School in Baltimore told the February FDA hearing that the commission "does not believe that these problems can be solved judiciously or expeditiously if tests are offered by private laboratories". The pilot project soon to

be launched with the cooperation of obstetricians and parent groups in the Baltimore area will be organised through the Laboratories Administration of the state's Department of Health and Mental Hygiene.

One advantage of a state-run project is that it can provide the close coordination between different individuals and institutions that is essential for the success of the screening test, but might otherwise be lacking in a more laissez-faire arrangement. "The solution (to unnecessary risks) will require the multiple agencies of the Public Health Service to cooperate, and PHS to cooperate with the private sector". Dr Foege of CDC wrote to the FDA, urging that a plan for such a policy "must be worked out before a kit or kits are marketed."

In Britain, such requirements have the advantage of being able to work through the National Health Service, and screening is now offered by a number of regional health authorities. In the US, where any moves towards "socialised medicine" are still strongly resisted, such tight coordination is a relatively new departure. "The need for quality control means that the various agencies will have to be part of a cooperative system of testing much broader than we are used to. The problems are difficult, but not, we think, impossible," one FDA official said last week.

FDA is expected to announce within the next few weeks the path that it proposes to take through this complex of issues. Its decision will be closely scrutinised; Mrs Bucholz of the Spina Bifida Association, for example, has warned that if restrictions on licences "are not carefully designed and planned for with due consideration we will not only withdraw our support of alpha-fetoprotein screening but will actively work to inhibit AFP testing".

On the other side, laboratories and manufacturers could well seek recourse to legal constraints if they feel FDA is imposing unwarranted restrictions on their trade. But in the current absence of restrictions, some small laboratories are already offering AFP screening on a "bootleg" basis, with minimal control over either the quality of their test results or the way they are interpreted and used.

Thus what was until recently a largely ethical debate on the social implications of genetic screening has become an overtly political issue. In the case of nuclear power — perhaps not strictly comparable, but one where the same type of issues about controlling risk have arisen — it took the Three Mile Island accident to awaken broad public debate on the dangers of leaving the "control of quality control" too firmly in the hands of private enterprise, with insufficient public accountability. "We hope that the FDA will not make the same mistake" one observer said last week. □

Handle us with care

Alastair Hay reports on the reaction of chemical companies in the UK and US to safety legislation and control of toxic substances

THE chemical industry is reacting in no uncertain manner to toxic substance control measures, introduced by both the US and UK governments as a result of recent public pressure and originating in concern over industrial pollution.

In the US in particular, the industry is objecting to proposals for testing the toxicity of new substances. Singled out for attack is the 1976 US Toxic Substances Control Act, which authorises the Environmental Protection Agency (EPA) to obtain production and test data from industry on selected chemical substances and mixtures, and to regulate the substances when needed.

The chemists are acutely aware of the political pressure to curb pollution, and their reaction to these proposals is not an outright condemnation of the philosophy of control, but a more subtle questioning of the validity of some of the measures designed to achieve it. Two issues are central to the chemical industry's concern: the variation found in the government control measures of different countries; and the scientific validity of testing procedures for toxic substances.

Britain's Health and Safety Executive (HSE) is sensitive to the protests from industry, but as the public guardian of the nation's safety, it is determined to achieve safer working conditions. The HSE policies the 1974 UK Health and Safety at Work Act, which requires industry — as the employer — to ensure that workplaces are safe. However, the means whereby this safety is assessed is still a matter of debate.

Last year the HSE issued a discussion document on a Notification Scheme for Toxic Substances, which would provide an agreed procedure for toxicity studies. There was no shortage of response, and this reaction has been taken into account by the Executive in the preparation of the next stage of the scheme: a consultative document. Views on this will again be sought before the notification scheme eventually becomes law.

The approach of the HSE to the regulation of toxic substances is in marked contrast to that of the EPA. Although the EPA guidelines for toxicity testing have yet to be finalised, one early draft suggested a specific battery of tests for assessing the toxicity of compounds. The HSE proposals on the other hand, are more flexible and selective. Mr Ted Langley, who has overall responsibility for the HSE notification scheme, outlined the philosophy behind the required toxicity testing as "the need to have a minimum set of data to enable movement to be made to the next stage (of development of a chemical) rather than a rigid set of tests, which may be very expensive." According to Langley the essence of the HSE approach is flexibility to "identify the hazard."

The HSE and EPA also differ in their approach to toxicity testing. The Executive's policy is based on an instalment plan, where sub-acute toxicity data generated over a 28 day study period will provide sufficient information to enable evaluations

to be made about a chemical's toxicity. If the data generated in this period suggest the need for further research of a long term nature the HSE will then insist that this be done. The EPA, on the other hand, envisages a one-off, all-embracing package, which requires that a battery of tests, both short and long term, be undertaken before a chemical will be considered for registration.

The Chemical Industries Association (CIA) is the collective voice of the British chemical industry with a membership ranging from the huge multinational companies like ICI to very small concerns. On the question of safety the Association is on record as "fully supporting" a notification scheme for new chemical substances, but not unnaturally, it has some reservations about the new HSE proposals and even graver ones about the EPA Toxic Substances Control Act.

Mr Roy Granger is Director of Company Operations for the CIA and, as such, responsible for communicating the Association's views on the new regulations to the HSE and EPA. Granger does not mince his words when commenting on their proposals. On the Toxic Substances Control Act he says the "EPA shopping list of tests is unjustified" and that the CIA does not "want a rigid programme of tests." Granger approves of the HSE proposals which he says are based on a dialogue between the Executive and industry. In Granger's view the HSE

proposals represent the current understanding of experts in the UK of the value of tests to assess toxicity.

In reply to the HSE and EPA proposals the CIA says it is concerned that UK proposals may differ from those being considered by the European Economic Community. It is also worried that the cost of doing the testing — and particularly the EPA tests — may prove so prohibitive for small or medium size chemical companies that it will be a barrier to innovation and to the introduction of new products on the market.

Furthermore, the CIA says that there is an acute shortage of toxicologists available to evaluate the tests; that intermediates in the chemical process should be exempt from legislation; that more thought must be given to the confidentiality of company data; and that the UK proposals which require notification for chemicals marketed in quantities above 1 tonne per annum are too stringent, and the figure should be raised to 10 tonnes. Raising this figure, says the CIA, would exempt many products and certainly most development chemicals from legislation.

For its part the HSE has let it be known that it is anxious to ensure uniformity in European legislation so that British manufacturers will not be unfairly penalised, and that HSE proposals may ultimately have to be modified to achieve these ends. The number of toxicologists available to evaluate data is not regarded as a long term



problem by the HSE, which points out that if necessary, it can look to Department of Health and Social Security toxicologists for advice. The HSE is building up its expertise in the field by an intensive "in house" training programme. The Executive will comment on intermediate products, confidentiality and tonnage in its consultative document.

The CIA claims in its submissions to the HSE and EPA that the views it expresses are those of its members. One such is Yorkshire Chemicals, a medium sized company based in Leeds, and dealing mainly in dyestuffs. It employs some 900 people and has an annual turnover of £22 million. Mr John Dawson, research director for Yorkshire Chemicals, agrees with most, although not all, of the CIA's views, saying for a start that the cost of toxicity testing is high. He feels that some of the new tests required by the HSE and EPA are "not justified" for dyes because, in his opinion, the tests required should be related to a product's final use.

To the best of his knowledge dyes themselves have never killed anybody, although he concedes that in the past his industry has used products such as β -naphthylamine and benzidine — which are carcinogenic in man — as intermediates. But he insists that toxic products such as these are no longer used.

Dawson's view is not shared by David Douglas, newly appointed head of the HSE unit responsible for collating information on toxic substances. Douglas insists that Dawson is "out of the mainstream of thinking on this one," as there is a risk to people from the breakdown products of finished dyes.

Many medium size chemical companies use outside research institutes, such as the Huntingdon Research Centre, to carry out toxicity testing. The dye stuff industry also shares much of its research data by operating collectively through the Ecological and Toxicological Association of Dyestuff Manufacturers (ETAD).

According to Dawson, the limited range of tests recommended by ETAD to assess the safety of their products are sufficient to assess the toxicity of dyes. In Dawson's view, "better hygiene could obviate the need" for some tests such as chronic oral dosing studies, and he doubts whether the CIA recommendation on raising the minimum tonnage of chemicals requiring testing would have a great effect on Yorkshire Chemicals.

Although many chemical companies complain about the cost of the HSE's proposed notification scheme, Ted Langley points out that the outcry is perhaps unnecessary. If, he says, industry is carrying out its obligations under the 1974 HSW Act to protect its workforce, then it should already be doing much of the testing which will be a feature of the new notification scheme. The difference, however, is that there will now be an agreed set of tests to do, and the results will have to be submitted to the Executive for evaluation. □

How ICI tests for toxicity

Few chemical companies can be as well prepared to undertake toxicity testing as ICI. The company finances its own laboratory, the Central Toxicology Laboratory (CTL), at Alderley Edge in Cheshire. CTL has a budget of £6 million for 1979. With a staff of nearly 300 the laboratory works in close collaboration with ICI's Occupational Medical Service (OMS) and both CTL and OMS have had substantial influence on the philosophy behind the HSE approach to chemical toxicity testing in the UK.

The science of toxicology is still in its infancy, according to scientists at CTL. They believe that research into new methods for assessing toxicity is essential to generate new tests which may be more rapid but, most important, are valid and reliable. Dr Iain Purchase is an Assistant Director at CTL and he believes that both the search for new test systems and work on understanding toxic actions are essential programmes for any such laboratory.

Such research, he believes, provides the right blend of academic work and routine testing which allows toxicologists to develop and execute an alternative to the recipe book approach to toxicology. Furthermore it provides the laboratory with a good understanding of the advantages and limitations of tests to assess toxicity.

It is this experience which makes CTL so influential. One area in which the laboratory has had considerable impact — and particularly on the HSE — is in the field of short term *in vitro* tests for predicting carcinogenicity. A cell transformation assay developed by Dr Jerry Styles at CTL is at least as reliable as the Ames test for detecting potential carcinogens, and in fact complements it. With some known carcinogens such as Butter Yellow the Ames test fails to identify the chemicals as potential carcinogens whereas the Styles assay will do so; and the reverse is also true. On this basis, the CTL group advocates that where short term tests are used to assess carcinogenicity, as a minimum requirement the two tests should be used as a screening net. The HSE is in agreement with this suggestion.

Styles and his colleague at CTL, Dr John Ashby, have recently argued that *in vitro* tests have distinct limitations and that the evidence they produce is not sufficient to allow quantitative assessments about carcinogenic potency. They are, however, ideal for use in a qualitative fashion for identification of potential carcinogens. The point was well taken by the HSE, which believes that carcinogenic potency is best assessed by long term animal studies.

With CTL bearing the responsibility for the testing of ICI industrial

chemicals, some of which may be toxic and potentially carcinogenic, rigorous laboratory hygiene is essential. To achieve this CTL uses a concept pioneered by ICI to assess safety on its industrial plants — the hazard and operability (haz-op) study. Purchase reports that the studies were time-consuming, but worthwhile. They involved a detailed assessment of every stage in the testing of a chemical from its synthesis to the analysis of animal tissue following long-term feeding tests. Flow charts are constructed detailing the stages and every conceivable mishap is considered and precautions instituted, either to prevent it occurring or to contain it (for example in the case of spillage of a carcinogen). Purchase believes that this kind of analysis ought to be carried out in all laboratories dealing with toxic substances, be they chemicals, viruses or bacteria. He says that the analysis ensures far more professionalism in the laboratory.

Things are more complicated when it comes to dealing with an ICI process which involves the use of chemicals which might be carcinogenic. Although this is not a new problem for ICI, it is one which is becoming more complex. To deal with it the company has instituted an additional reviewing step in the form of a carcinogen panel. With a membership embracing the four disciplines of medicine, biology, chemistry and law, the panel meets fortnightly and since its inception in 1975 has reviewed some 150 chemicals. The panel's brief is to review all relevant data, from the literature and within ICI, and to advise on the chemicals which may pose a risk and the action that is required to deal with it.

Purchase says that the philosophy behind the actions of ICI and much of the chemical industry is that "hazards can be contained". Dr Eric Longstaff, one of the panel members, says that in 9 out of 10 cases it is an ICI division which approaches the panel for information and advice. If the carcinogenicity of a chemical is established and it is involved in an ICI process then the panel's review will probably ensure containment of the chemical.

Work at CTL embraces a far wider field than simply testing for carcinogenicity. CTL's toxicologists recognise that the need for good toxicity testing on chemicals is a fact of life, and while they admit that the ICIs of this world have the skills available to perform the tests required by Governments, their real concern is whether they have the capacity to cope with the work-load the testing will involve. The outstanding question is whether the smaller chemical companies, with fewer resources, will be able to cope too. □

correspondence

How accurate was the carat?

SIR, — The carat, a well known jeweller's measure of weight, is generally defined as the weight of a seed of the carob tree (*Ceratonia siliqua*; Leguminosae) or about 200 mg (*Webster's New World Dictionary*). It is traditional to view seeds as having very low intra specific variation in weight, and therefore not surprising to find them used as standards of commercial weight. But how reliable was this species of seed as a standard?

The carob is a small tree native to North Africa and has the largest dry and hard seed of any Mediterranean plant. In September 1978, I haphazardly collected 12 freshly fallen pods from the ground below a healthy carob tree growing in a courtyard of the Hebrew University, Jerusalem. In April 1979, the seeds were removed from the indehiscent pods and weighed. There were ten seeds that weighed only 0.07 to 0.13 g, and I discarded these seeds from further analysis because they were variously malformed and would be noticed by any discerning merchant or shopper. The remaining 156 seeds could not be sorted into groups of heavier or lighter seeds, and this set had a mean weight of 0.198 g (s.d. = 0.024), remarkably close to the supposed weight of the carat.

However, there is no reason to believe a merchant would be so stupid as to use average (randomly selected) carob seeds as balance weights. There were 48 seeds (31%) in the sample that looked normal but weighed less than 0.2 g. All a merchant would have to do is use a gravitational sorting system on a large sample of seeds to obtain a pool of carob seeds up to 25% underweight to use when making sales. The other end of the weight distribution contained 64 (41%) that weighed 0.21 to 0.24 g. These heavy seeds would have been most useful when making purchases. In short, only 28% of the seeds in a carob seed crop would have given an honest measure. This study was supported by a subsidy from the Israeli government.

Yours faithfully,

DANIEL H. JANZEN
University of Pennsylvania, Philadelphia,
Pennsylvania, USA.

Science is not all whizz

SIR, — Few would disagree with your leader article (10 May, page 89) that science in the UK is depressed. More precisely, scientists who have chosen to make a career in research are depressed. The danger is that the full effects of the lack of career structure for research scientists has not yet made sufficient impact on scientific output in this country to force any action on the problem. By the time it does, it may be too late. The experienced scientists who should play a major part in research teams of the future will have been dropped by the wayside because research councils and charities are no longer prepared to take on the "burden" (as they describe it) of funding the older person.

Your challenge to the new government to find effective, creative employment for Britain's postdoctoral fellows is to be welcomed. But is it centres for "whizz-kids" that we really need? Advancement in scientific research rarely requires the flash of genius, but

rather a methodical and pains-taking approach. Most research scientists are ordinary men and women of average intellect trying to do a job like any other professional. How good they are at their job depends critically on the kind of training they have received. It is about time we got away from the notion that only geniuses can produce good science, and the corollary, that only scientists of "exceptional ability" should be funded. We are not arguing for the support of mediocrity, merely the acknowledgement that scientific research, like other professional careers, should have room for a wide variety of people with differing abilities and contributions to make, not just jobs for "whizz-kids".

Yours faithfully,

SUSAN BARLOW
ARMS, Guy's Hospital, London

Physics and Ghana

SIR, — Your article (10 May, pages 96-97) rightly emphasised the relative irrelevance of electron transport in thin manganese films to scientific problems in Ghana.

However, as the supervisor of the physicist concerned during his studies, partly in Ghana

and partly at Sussex on a faculty exchange, I would like to add that we also carried out an extensive literature search and correspondence with the Timber Research Association relating to possible physics-based problems in Ghanaian forestry. These formed the basis of many long discussions during his time at Sussex.

Yours faithfully,

A. D. C. GRASSIE
University of Sussex, UK.

DPAG, GMAG and the trade unions

SIR, — Reg Bird (26 April, page 776) uses the argument that, because a fatal accident occurred under the supervision of DPAG, which has no trades' union nominees, whereas no comparable accident has occurred under GMAG, which has TUC nominees, therefore the accident free record of GMAG is a consequence of this presence. It seems a pity that fallacious reasoning which would never appear in your scientific papers is allowed to exist in your correspondence columns.

Yours faithfully,

ALAN D. B. MALCOLM
St Mary's Hospital Medical School,
London, UK.

Nuclear power does not lead to nuclear weapons

SIR, — Professor Rotblat with his phrase "the life blood of modern society" (31 May, page 370), recognises the need for adequate energy supplies, yet he seems reluctant to accept that nuclear power should make a contribution.

Even though his proposed World Energy Organisation would (*inter alia*) "advise countries of the type of energy most suitable for them — including nuclear if deemed necessary", Professor Rotblat's obsession with the proliferation of nuclear weapons, the extreme views he put forward in his lecture as reported in *Nature*, 1 March, page 4 ("In the long run nuclear energy is not compatible with the survival of civilisation") and his hostility towards the IAEA make it clear that nuclear power would play little part in his WEO.

Although my arguments (26 April, page 776) have failed to convince Professor Rotblat, the facts seem plain. There is no case where a country has developed nuclear explosives by diverting material from a civil power station. In the ten year period 1945-54 three countries (USA, USSR and UK) developed nuclear explosives; between 1955 and 1964 another two countries (France and China) and between 1965 and 1974 one further country (India). Over this same time span the world nuclear generating capacity grew, and by 1974 there were 54,000 MW in operation in 19 countries. There are now 100,000 MW in operation in 22 countries, and if stations under construction are also included, the figures rise to 400,000 MW in 33

countries. This indicates that there is no direct link between peaceful nuclear power and nuclear weapons. This point comes out even more clearly when it can be shown that for each of the five nuclear weapon states the explosion of their first bomb preceded, not followed, the entry into service of their first prototype commercial power reactors (Table 1).

The plutonium for the Indian nuclear explosion of 1974 was obtained from a research reactor. In addition it is widely believed that Israel and South Africa are probably capable of making nuclear weapons, but neither country has a commercial nuclear power station in operation.

It could be argued that the greater a country's dependence on nuclear energy for essential electricity the less likely it would be to jeopardise its relations with international suppliers by breaking international agreements. Quite contrary to the innuendo of Professor Rotblat's final sentence, most nations have not become nuclear weapon states: 104 of them have acceded to the Non-Proliferation Treaty.

It is essential that every effort is made to find ways to prevent the misuse of nuclear energy; a refusal to accept its benefits is not one of them.

Yours faithfully,

G. H. GREENHALGH
A Power for Good,
London, UK.

Table 1 Dates of first nuclear bomb and first nuclear reactor for the five nuclear weapon states

Country	Date of first bomb exploded in atmosphere	Date of first prototype power reactor in service	Commercial name
USA.....	1945	1957	Shippingport
USSR.....	1949	1958	Troitsk
UK.....	1952	1956	Calder Hall
France.....	1960	1964	Chinon
China.....	1964	No firm evidence of a power reactor in service	

news and views

The legacy of Xenophon

from Robert W. Cahn

THE US National Academy of Sciences gave the name COSMAT to a massive study it sponsored of the profession of materials science and engineering (MSE), published in five thick volumes in 1974 and 1975. It is a remarkable compilation; because of its wide terms of reference, and especially because of its historical and international emphasis, it is of value far beyond the originally intended readership in its target profession. The report is difficult to locate and few people in Britain have read it, and to make at least a small extract accessible to a wider readership, a learned periodical, appropriately named *Materials Science and Engineering*, has published a special issue entirely devoted to a revised and updated version of two chapters from one of the volumes (*Materials Science and Engineering: Its Evolution, Practice and Prospects*, Elsevier-Sequoia, March 1979).

The heart of the extract is a forty-page essay, by Melvin Kranzberg (editor of the journal *Technology and Culture*) and Cyril Smith (the world's leading scholar in the history of metallurgy), on 'Materials in History and Society', complete with a bibliography of 85 items which alone would justify acquisition of the volume. The emphasis is predominantly on the history of the use and understanding of metals. The essay is a masterly examination of the ways developments in the manufacture, purification, variegation and fitness for purpose of materials influenced, not only the tenor of daily life, but matters of great social impact such as the spread of agriculture; the balance of affluence between patricians and plebians (as 'aristocratic' bronze gave way to 'democratic' iron); the spread of the printed word and the printed image; the materials testing institutions which were a precondition of safe machinery and all that implies. The later parts of the essay are more concerned with the gradual development — remarkably delayed, and significant only in the twentieth century — of a scientific and analytical approach to materials. Réaumur, in the early eighteenth century, was the first modern materials scientist, but his ability to relate traditional practice to advanced science was not followed, for the love of

mathematical physics applied to highly simplified models was at that time irresistible and the obscurantism of the alchemists had repelled the disciples of Newton.

Kranzberg and Smith's survey is followed by a section with its eyes wholly on modern times: Richard Claassen and Alan Chynoweth, two industrial scientists, write on 'MSE as a Multi-discipline'. Specifically, they trace the routes by which physicists, chemists and engineers of various persuasions combined to create the profession of MSE, and they do so primarily through a collection of case histories of technical developments (heatshield design for space vehicles; the development of the transistor; plastic coatings on razor blades; artificial fibres; optical lightguides, and others). These stories, combined with an analysis of the implications for the organisation of interdisciplinary research projects, provide the flavour of this section.

Kranzberg and Smith, the historians, quote the Greek historian Xenophon ("What are called the mechanical arts carry a social stigma and are rightly dishonoured in our cities") and Plutarch ("For it does not of necessity follow that, if the work delight you with its grace, the one who wrought it is worthy of esteem"), to indicate what the ancients thought of craft and skills applied to materials. Of all the ancient exponents of the 'mechanical arts', the slave miners of antiquity were perhaps the most abjectly oppressed, miserable and shortlived.

Today, once again, 'mechanical artists' and miners in particular are apt to be regarded by wielders of temporal power as the carriers of social stigma. This surprising conclusion emerges with great clarity from another recent publication, the proceedings of the 1978 American Mining Symposium, the subtitle of which was 'Politics, Society and the Mineral Industry'. The proceedings are published in *Materials and Society* (3, 1; 1979). Several of the lectures are germane to the matter in hand: these are written by a metallurgist, a congressman, a copper mining expert, a coal mining expert and J. Allen Overton, the president of the American Mining Congress, who finds

himself under the necessity of acting as a fulltime lobbyist. Their complaint — the word is too weak, it is a succession of cries of despair — is the unchecked growth of regulation and of the body of state employees who enforce regulations and steadily add to their number. The chairman remarked that the 'political outcry' as he termed it, at the symposium was unpremeditated and uncolluded and the more credible for it. The tone is exemplified by two extracts from Allen Overton's lecture: "Time and again, we see no attempts to consider trade-offs or weigh the cost of regulations against their supposed benefits. This is the work of zealots, and the only guideline they know is their own self-proclaimed virtue. The repeated failure — I should call it the obstinate refusal — to take account of the economic consequences mean that some regulations can have ruinous impact." And again: "Finally, there is bewilderment and foreboding about the sheer mass of regulations, their frightening complexity, the paralysing uncertainty about what the bureaucrats will propose tomorrow, and how the rules will be changed in the middle of the game". The changing of rules in the middle of the game is exemplified by William Simon, a former US Secretary of the Treasury, in another recent publication (*A Time for Truth*, McGraw-Hill, 1978): "Manufacturers of children's sleepwear were forced by the government to process children's sleepwear with a flame-retardant chemical. When the companies shifted over to the costly new process, the FDA banned the flame-retardant chemical as a suspected carcinogen, and the manufacturers were ordered to recall their merchandise and to compensate buyers".

In such an atmosphere, it becomes understandable why American oil policy is in such disarray: those who fully understand the situation in prospect plead for decontrol and a partial release from the burden of regulation (McKetta's lecture on American energy policy at the symposium is a case in point), but their

Robert W. Cahn is Professor of Materials Science at the University of Sussex.

word is widely distrusted by politicians precisely because they are experts. Xenophon stalks among us again. The proponents of regulation, and those who support and encourage them, are all too often convinced that experts are venal and devoid of public spirit (just as the free citizens of ancient Greece regarded their bonded craftsmen). When the mining specialists plead public interest, the fact that the plea comes from them at once invalidates it. But the inevitable consequence is that the same judgment is being applied to the regulators themselves: they have by now as great an

interest vested in the increase of regulations as the protesters have in their abatement. Not only that: those miners, MSEs, and many others to whom by implication Xenophon's cold dictum is attached, conclude that their honesty is disbelieved by some of the most powerful in the land, and when a professional's integrity is systematically impugned, he is driven at last towards political extremes; at the least he will come to distrust the regulators' own honesty of purpose. The prospect is not enticing, and the American crisis of regulations should give us in Britain pause. □

recessive scrapie allele is likely to be widespread but clinically 'silent' in these breeds and may at any time without prior evidence reach frequencies leading to clinical manifestation.² Thus the group designated 'scrapie-free' is likely to consist of mixed populations of the three genotypes. The authors quoted by Kimberlin provide no indication that the proper safeguards against these possibilities have been enforced. The meaningful interpretation of results from such populations on the present evidence is thus not possible.

The second consideration is that the results quoted are commonly presented as totals of all matings between so-called 'scrapie-free' and 'scrapie-affected' groups without any attempt to assess presumptive recessive genotype frequencies in each group. Third, the aggregation of the results of different sire-progeny groups as accumulated totals masks information relevant to their proper evaluation, for example, the possible proportions of the three recessive genotypes and allele frequencies in different sectors of the breeding population, the degree of in-breeding, occurrences of associated sub-normal health and reproductive efficiency, which may affect the totals manifesting scrapie and their ratios. Fourth, the American experience, quoted by Kimberlin, kindly made available by Dr. J. L. Hourigan, when analysed, as far as the data allow, on a provisional genotype basis, are approximately compatible with my own. I therefore have serious reservations of the validity of the conclusions derived from the experimental flocks quoted as evidence for the normal dissemination of natural scrapie by the spread of a communicable agent.

Artificial scrapie and genetic control

The artificially induced form of scrapie disease is probably not identical to the natural form and the precise transferring of the results of the one to the other is questionable. In most of the studies quoted by Kimberlin the source of inocula of scrapie TSEPA material designated SSPB/1 was derived from 40 affected sheep brains. After 20 years this is now very much more toxic than any TSEPA we have been able to demonstrate in the natural disease on first transfer.

These reservations do not preclude extrinsic factors operating in natural scrapie, merely that their possible occurrence requires more precise definition. Any aetiological hypothesis must also provide an acceptable explanation of the specific characteristics of the neuronal degeneration in the natural disease.⁸ □

H. B. Parry works at the Nuffield Institute for Medical Research, University of Oxford, UK.

Aetiology of natural scrapie

from H. B. Parry

THE crux of the problem raised by Kimberlin recently in *News and Views*¹ is the principal means by which natural scrapie of sheep is disseminated. My report², which stimulated him, was confined to providing new evidence for a hereditary factor and to demonstrating a new practical means of genetic control under current agricultural practice. Kimberlin's account,¹ summarising the published laboratory studies of his colleagues,³ widens the issue of aetiology without considering relevant evidence from the epidemiology of the natural disease or from the possible pathogenesis of the characteristic sequential localised neuronal decay.

There are two main aetiological factors. First, a neurotoxic agent is present in the tissues of animals with the pathological lesions of scrapie and is transmissible artificially.⁴ This is commonly called a 'slow virus', or more accurately a transmissible spongiform encephalopathic agent (TSEPA) from the pathological nature of the special neurotoxic damage developing in inoculated animals. In the absence of any reliable means of detecting the TSEPA in the living animal, we have no data on the occurrence of the agent in any sheep population. Second, there is a hereditary factor, probably a single autosomal recessive gene, with an 'all-or-none' effect on the development of the primary neuronal decay, for which no simple peripheral 'marker' reaction is known. However, indirect inferences of the gene's likely presence can be made on the basis of the results of test-mating procedures and meticulous breeding records.¹

Without such data, assessments of recessive allele frequencies are notoriously hazardous when based solely on the statements of individual flockowners and shepherds.⁵⁻⁷ In my view the term 'scrapie-free' should mean free of the scrapie-allele, at least to a probability of less than

1 in 100; such sheep are available.

Spread of scrapie

The inferences of natural spread among experimental flocks, quoted by Kimberlin, are contrary to my field experience within breeding groups. Using recording methods designed to reveal such spread, I detected none, which is in accord with the evidence of many reliable observers since 1750⁵⁻⁷ and with the results of other less well controlled current field observations. No spread has followed the introduction since 1950 of British sheep, which developed the disease, into certain countries that are free of the disease and have not operated any quarantine or slaughter policy. One therefore seeks possible explanations for the anomalous conclusions quoted by Kimberlin. There are four main considerations.

First, the data he quotes relate to four British sheep breeds, all of which have had substantial scrapie attack-rates at some period since 1900 and three very seriously in some sectors of their breeds since 1950 until the present time. In these circumstances the designation 'scrapie-free' applied to a sheep, a flock or a breed without meticulously supervised and relevant records, covering at least one decade before and continuously during the experimental period, lacks any precise scientific meaning. This is because the TSEPA cannot be detected and the

1. Kimberlin, R. H. *Nature*, **278**, 303 (1979).

2. Parry, H. B. *Nature*, **277**, 127 (1979).

3. Kimberlin, R. H. (ed) *Slow Virus of Diseases of Animals and Man* (North-Holland, Amsterdam, 1976).

4. Oppenheimer, R. H. in *Viral Diseases of the Nervous System* (ed. Illis, L. S.), 161 (Bailliere Tindall, London 1975).

5. Fink, J. W. *Communications to the Board of Agriculture (London)*, **1**, 276 (Nicol, London, 1797).

6. May, G. *Das Schaf, Seine Wolle, Rassen, Zucht, Ernährung u. Benutzung, sowie dessen Krankheiten* **2**, 242 (Trewendt, Breslau, 1868).

7. McGowan, J. P. *Investigation into the Disease of Sheep called 'Scrapie'*, 20, 101. (Blackwood, Edinburgh, 1914).

8. Parry, H. B. & Livett, B. G. *Neuroscience*, **2**, 53 (1977).

Photogenic impurities

from Henry Slayter

Two papers, published together in the *Journal of Biological Chemistry* (254, 1747; 1979) come essentially to the same conclusion concerning the structure of pyruvate carboxylase examined by electron microscopy. Goss *et al.* of the University of Adelaide and Cohen *et al.* of Case Western Reserve and the National Institute for Medical Research, London, have both applied negative staining to this molecule, obtaining different results from those previously reported by Valentine *et al.* (*Biochemistry* 5, 3111; 1966) using the same technique. (Two of the coauthors of the latter group were also coauthors of the original Valentine group.)

Negative staining was first described and applied by C.E. Hall and later applied skillfully and to great advantage by many investigators (including, especially, Valentine and R.W. Horne) in the study of biopolymeric structures. In this method, the purified preparation is mixed with concentrated electron-dense 'stain', usually a 1–5% solution of uranyl acetate, uranyl formate, sodium phosphotungstate or sodium phosphomolybdate. The preparation is air-dried to form a thin layer supported on 10-nm carbon films. Stain, which is excluded by biopolymer components, penetrates voids, thus outlining the structure. Electron contrast results from the greater electron scattering power of the stain (average density about 7) compared with that of the specimen.

Inherently, this technique is a simple and effective method for examining macromolecules larger than about 5 nm in diameter — but pitfalls abound. Resolution is limited by granularity of the stain, alteration of specimen structure by dehydration, extent to which contrast can be induced in small subunits, and, to an indeterminate extent, specimen damage in the electron beam. Other difficulties relate to the effects of these rather unusual chemical salts on the specimen (due to pH and ionic strength requirements, the reducing nature of the salts, the presence of divalent cations and so on). Penetration of the specimen by stain can produce misleading measurements of dimensions or loss of fine structure. As in any method for the examination of individual macromolecules, absolute specimen purity is vital.

Valentine *et al.* originally described the pyruvate carboxylase molecule as a distinctive square-planar tetramer of globular subunits. Several lines of evidence supported this premise. The highly photogenic square-planar configuration appeared in enzyme preparations isolated from several animal species. Dimensions

obtained by electron microscopy were consistent with the measured molecular weight of 500,000, consisting of four similar or identical subunits of molecular weight 125,000. The square-tetramers exhibited cold-lability, a well characterised property of chicken liver pyruvate carboxylase.

Valentine, however, also showed that the same enzyme from yeast had a rhombic appearance (Valentine, *Proc. fourth Regular Conf. (Europe) on Electron Microscopy, Rome 2, 3*; Tipografia Poliglotta Vaticana, Rome, 1968). Now, the new reports show that a similar structure characterises pyruvate carboxylases of animal origin, and that the square-planar molecule is a contaminant commonly found in such preparations. That this contaminant was so much more evident than the enzyme itself is explained, in part, by the demonstration of Goss *et al.* that this structure is actually an octamer in which the thickness (and therefore the relative contrast) of each subunit is doubled.

Both new studies confirm that pyruvate carboxylase consists of four subunits. Cohen *et al.* suggest a rhombic structure with major and minor dimensions of 19 nm and 16 nm respectively, and a centre-to-centre distance between adjacent subunits of 7.3 nm. (They report that exact dimensions vary by as much as 10 % depending on the substance used as negative stain.) Goss *et al.* describe the molecule as a 'splayed tetrahedron' with a long axis of approximately 17.7 nm and a short axis of 15.2 nm with centre-to-centre distance again 7.3 nm. The main difference between the two interpretations is whether the subunits lie in a plane (rhombic structure) or are without regular mutual separation and are not coplanar (splayed tetrahedron).

Cohen *et al.* indicate that their data cannot be fully interpreted with regard to the question of subunit arrangement in three dimensions, but Goss *et al.* suggest that one pair of subunits is displaced slightly above the plane of the second pair. Examination of the field of negatively stained molecules presented by the latter authors shows that only a small number of images can be so interpreted. Thus, while their speculative model is most interesting, I must admit difficulty in accepting with certainty either the proposed three-dimensional arrangement or the detailed tear-drop model of subunits. Further understanding of the structure of this enzyme must await preparation of crystals and their examination by iterative electron microscopy or X-ray crystallography.

Anti-biotin antibody and pyruvate carboxylase are known to form soluble

complexes. Cohen *et al.* also present micrographs of this complex, which indicate that the antibody binds to the surface of the tetramer at the same general location as does avidin. Thus, additional evidence is presented that the rhombic tetramers are in fact a biotin-containing protein.

In perspective: some will remember that similar interpretive difficulties have arisen in past studies of negatively stained enzyme preparations — brought on, likewise, by impurities of a significantly more photogenic character than the plainer principal subject. (DNA-dependent RNA polymerase, first reported to be amorphous globular protein, was credited by a second group with a distinctive polyhedral structure. A third group showed up the polyhedron as a mere contaminant and found the polymerase to be an amorphous globular protein.)

What can be learned from all this — perhaps — is that while carefully applied electron microscopy can be of tremendous and unique importance in providing structural information, interpretation must always be carried out with extraordinary care. Awareness of the artefacts to which biopolymers are prone is a must, and investigations should include parallel studies in solution which pose the same questions asked of the electron microscope. Above all, very careful attention must be given to the purity of the preparation. Even a few per cent of contaminating molecules, especially those of the more photogenic sort, can — given the harsh treatment inherent in this method — be interpreted as being a few per cent of survivors of the structure sought. □

Budgeting for volcanoes

from Peter J. Smith

WHAT is the energy budget of a volcano? Or to be more explicit, what proportions of a volcano's energy are carried away by the various modes of energy transport? What, for that matter, are the appropriate transmission mechanisms? These are the sort of conceptually simple questions one would expect some nineteenth century volcanologist to have answered, albeit incorrectly; but whether he did or not, the same questions are still being asked in 1979, possibly because they are not as easy to come to terms with as might be supposed. Certainly they are not answered entirely satisfactorily by McGetchin and Chouet (*Geophys. Res. Letts.* 6, 317; 1979) even now, as these authors freely admit.

But to start at the beginning, for many years McGetchin and Chouet have been studying the behaviour of Stromboli, the famous volcanic island that rises about 920m above the Mediterranean some 75km

Henry S. Slayter works at the Farber Cancer Institute and the Department of Medicine, Harvard University.

north of Sicily. Since records began this volcano has been in a 'more or less continuous state of episodic activity, characterized by well-collimated jets of incandescent gas loaded with molten fragments of basaltic vitric lapilli' — a characteristic 'fire-fountaining' type of activity that has come to be known as Strombolian though not limited to Stromboli. In jargon, this is known as the volcano's 'normal mode', the 'steady state episodic mode'; and it is with this type of behaviour that McGetchin and Chouet are chiefly concerned.

No reliable heat flow values have been obtained in the vicinity of Stromboli; but from their own downhole temperature measurements, McGetchin and Chouet estimate that above sea level the volcano is emitting something like 6 MW in the form of heat conducted up through the ground. This is about 84 % of the total energy emitted. By comparison, all other forms of energy transmission are small and most are negligible. Next in magnitude comes the heat carried by ejected gas (700 kW, or 9.8 %), then the heat carried by ejected particles (2.8 %) and the heat radiated from the vent (also 2.8%). Seismic energy released amounts to 0.42 % of the total. Finally there is the kinetic energy of ejected gas (0.0091 %), acoustic power (0.0014 %) and the kinetic energy of ejected particles (0.00034 %).

Needless to say, these estimates depend on assumptions and are therefore open to correction. But even if the results are perfectly accurate, the conclusions that McGetchin and Chouet reach are highly dubious, as they themselves note. The most significant phrase in the last paragraph is probably 'above sea level'. The fact is that Stromboli is largely submarine; yet the underwater region has been ignored, and with it the large amount of heat likely to be transferred by the convective power of circulating groundwater.

The question of the 'normal mode' of volcanic activity is also crucial. It is well known that from time to time Stromboli departs from its usual behaviour and undergoes violently explosive eruption. During the explosive eruption of September 1930, for example, the mass of material ejected was equivalent to that produced during 300 years of normal activity. These violent phases, though unusual in the duration sense, strongly influence the energy/mass equation. So should they be included in the sums, or ignored on the grounds that they are atypical?

On the face of it, McGetchin and Chouet seem not to have achieved very much so far; their figures are, at best, misleading and in any case refer only to one type of

behaviour from one particular volcano at one particular period. On the other hand, they have gone some way towards the solution of a fundamental scientific problem, demonstrating in the process that the simplest questions to formulate are not necessarily the easiest to answer. The issue may also be of practical importance to the extent that volcanic heat is being investigated as a possible source of geothermal energy. □

Assessing effects of marine pollution

from B. L. Bayne

STUDIES of the chemical and biological aspects of pollution should go hand in hand. The chemical assessment of estuarine and marine pollution involves the analysis of water, sediments and biota to determine the levels of contaminants in the ecosystem. Because of technical difficulties of analysing water and sediments (Phillips *Env. Pollut.* 13, 281; 1977), chemical monitoring is usually concentrated on the biota to provide 'averaged values for the biologically available contaminants' (Goldberg *et al. Env. Conserv.* 5, 1; 1978). The biologist is therefore given the dual responsibility of elucidating the effects, if any, of the observed levels of contaminants in living material, and co-ordinating his efforts with those of the chemist in order to provide an overall assessment of environmental quality.

It is not surprising, therefore, that there has been a recent increase in interest in procedures for measuring the biological impact of pollution (International Council for the Exploration of the Seas (ICES) Cooperative Research Report 75, 1978; *Phil. Trans. R. Soc., Lond.* 286, in the press; ICES Workshop on Pollution Effects Monitoring, in the press). Much of the considerable research effort is based on the effects of pollution on the individual organism; from this work two general points have emerged which facilitate the planning of further research.

First, the non-specific effects of pollution can generally be characterised in terms of a stress syndrome following the paradigm of the general adaptation syndrome of Selye (*Stress*; Acta Inc., Montreal, 1950). This syndrome includes physiological (mostly involving energy gains and losses, and hormone balance), pathological and biochemical responses, the details of which may vary between species, but which, taken together, make possible a quantitative assessment of animal health that can be used to rank different populations according to the biological impact of environmental

changes (Bayne *et al. Phil. Trans. R. Soc., Lond.* 286, in the press). It is axiomatic that the observed biological responses must constitute reduction in ecological fitness for a pollution effect to be registered.

The second general point to emerge from recent studies is the widespread occurrence of mechanisms of detoxification in marine (and other) organisms. These include the cytochrome P450-linked mixed function oxidation reactions (Bard and James *Biochem. biophys. Perspec. mar. Biol.* 4, 126; 1978), the production of metal-binding proteins, or metallothioneins (Brown *et al. Env. Conserv.* 4, 213; 1977) and various cytological mechanisms, such as lysosomal encapsulation, that sequester some foreign compounds (Allison in *Lysosomes in Biology and Pathology*, ed. Dingle and Fell 2, 178; Elsevier, Amsterdam 1969). These are specific responses to particular classes of contaminant and they can provide evidence of the causative agents involved in situations where measurements of the general stress effects indicate a decline in animal health. The concept of linked non-specific and specific effects of environmental agents is consistent with attributes of the general adaptation syndrome.

It has therefore become possible, by making certain measurements of the general stress response of animals, and by linking these on the one hand to measures of potential detoxification processes and on the other to the results of chemical analyses of the body burden of contaminant, to assess the impact of pollution on the organism. However, the transition from laboratory demonstrations of this kind to practical involvement in environmental management is complicated by the problem of biological variability. The response of the individual organism is seldom a simple function of the intensity of the pollution 'stimulus', because the reactivity of the animal may vary, depending on a combination of intrinsic and extrinsic factors. This inherent variability must at present be accommodated in sample design, by optimising the sampling strategy over time and space.

The merit of such studies on individuals, however, is the early warning that can be signalled concerning environmental impact; individuals respond to sub-lethal stressors long before the effects are apparent in the population or ecosystem. Nevertheless, there is an urgent need also for studies on pollution effects at the level of the ecosystem, if

only because the problems of variability at the level of organism or population may thereby be reduced (Mann and Clarke *J. Fish. Res. Bd Can.* **35**, 791; 1978). Attributes of ecosystem function that may warn of impending damage are few at present. Indices of diversity and species richness, although controversial, have been effective in some circumstances (McIntyre in *Ecology of Marine Benthos*, ed. Coull, 301; University of South Carolina Press, Columbia, 1977; Sanders *J. Fish. Res. Bd Can.* **35**, 717; 1978.). Greve and Parsons (*Helgol. Wiss. Meeresunters.* **30**, 666; 1977) have discussed the implications of a change in the ratio of diatoms to flagellates as an index of pollution effect, and Gray (*Phil. Trans. R. Soc., Lond.* **286**, in the press) has postulated that changes in the form of the cumulative species distribution curve indicate the effects of environmental disturbance. Measures of the productivity of various trophic levels, and of various size-fractions of the fauna, in both the pelagic and the benthic components of the ecosystem, may also be useful.

Further research is urgently needed and can be framed in certain generalised notions of biological response to stress. When this is coupled with better chemical understanding of the form and the biological availability of pollutants, it should be possible to provide integrated measures of environmental quality as a basis for legislation. □

Alpha-particle mode dominates giant resonance decay

from P. E. Hodgson

A GROUP at the National Bureau of Standards in Washington DC has recently made an important advance in an old technique, the use of virtual-photon spectra to determine the multipolarity of nuclear excitations (Wolynec, Dodge & Hayward *Phys. Rev. Lett.* **42**, 27; 1979). This has been used to study the nickel isotopes, and has led to the surprising conclusion that alpha-particle emission is the dominant decay mode of the E2 isoscalar giant resonance.

It is well known that when electrons are decelerated, for example when they are deflected by passing through a nuclear field, they produce photons, a process known as bremsstrahlung. These photons may be absorbed by the nucleus itself, initiating a nuclear reaction. In this case the photons are not emitted, and are referred to as virtual photons. The theory of these

processes is well understood, and it is possible to calculate the virtual photon spectra associated with the excitation of the E1 and E2 giant resonances, and hence to determine the energies and strengths of these resonances from the measured yield curves. The accuracy of these analyses has now been increased greatly by the use of the distorted wave Born approximation to calculate the spectra of the virtual photons.

In the new experiments, targets of the nickel isotopes ^{58}Ni , ^{60}Ni and ^{62}Ni were bombarded with electrons from 16 to 50 MeV, and the yields of protons and alpha-particles were measured at 48° , 90° and 132° . The spectra of these decay particles peak at 5 MeV for protons and at 8 MeV for alpha-particles, and have the asymmetrical shape characteristic of emission at energies around that of the Coulomb barrier.

The total cross-section for proton emission as a function of electron energy is shown in Fig. 1, and it is compared with calculations using the E1 virtual photon spectrum and the (γ, p) cross-section known from another experiment. The excellent agreement between the curve A and the experimental data supports the correctness of the analysis.

In the second part of the experiment, a tantalum foil was interposed in the electron beam in front of the nickel target. The photons emitted from bremsstrahlung in the tantalum then cause photo-disintegration of the nickel, so that the yield from these reactions is added to the electrodisintegration yield. The cross-section for the emission of protons with this foil in place is also shown in Fig. 1 (curve B), and agrees with the calculated sum of the electrodisintegration and photodisintegration cross-sections.

The cross-section for the emission of alpha-particles was measured in the same way, and the results are shown in Fig. 2. As

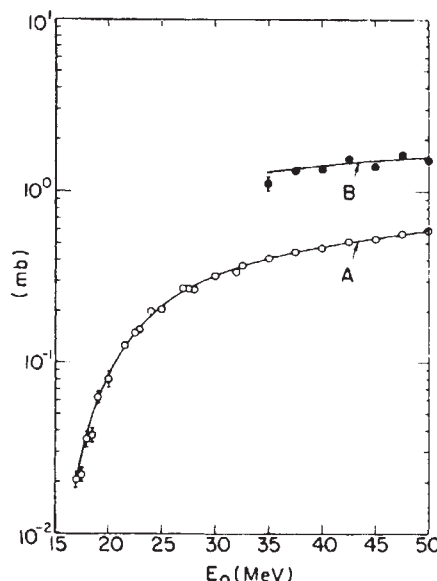


Fig. 1 Total cross-section for the production of protons as a function of the energy of the electrons incident on a ^{58}Ni target. The curve marked A refers to the case without, and B to the case with, a tantalum foil.

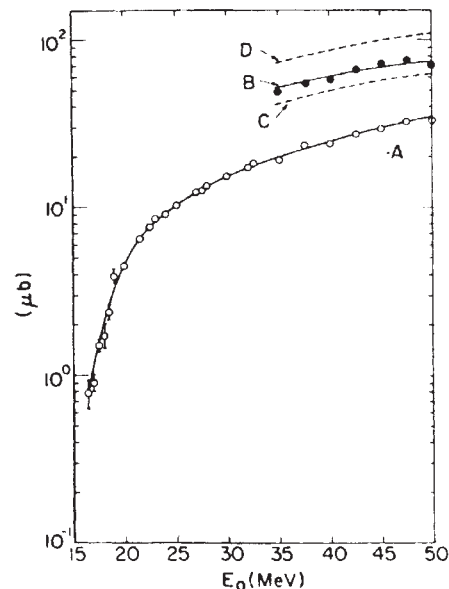


Fig. 2 Total cross-section for the production of alpha-particles as a function of the energy of the electrons incident on a ^{58}Ni target. The two sets of results are as in Fig. 1, and the assumptions made in obtaining the curves B, C and D are described in the text.

before, curve A refers to the measurement without the tantalum foil, and curve B to that with the foil. In the latter case there are three calculated curves: B was obtained using the E1 and E2 virtual photon spectra, C using the best E2 fit and D with the best E1 fit. Of these curves, only B agrees with the data, showing that both the E1 and E2 components must be included to fit the cross-section for alpha-particle emission.

These analyses show that the E1 components in the (γ, α) reaction have an energy of about 20 MeV and width of 6 MeV and account for only about 1% of the E1 absorption cross-section for each of the nickel isotopes. The E2 components have similar energies and widths, 16 MeV and 4 MeV, but account for more than half the energy-weighted isoscalar sum rule for ^{58}Ni and ^{60}Ni , and nearly a third for ^{62}Ni . Comparing with the values of 55% and 63% obtained from inelastic alpha-particle scattering for ^{58}Ni and ^{60}Ni , this leads to the conclusion that alpha-particle emission is the dominant decay mode for the E2 isoscalar resonance.

This is a very striking result, and it suggests that the E2 giant resonance has in some sense an alpha-particle structure. One possibility is that it consists of an alpha-particle coupled to a nuclear core. To test such a possibility, it will be necessary to make detailed calculations to see if it is in agreement with the experimental data. This latest application of the virtual photon method will certainly stimulate further studies to improve its accuracy and to apply it to other nuclei. □

P. E. Hodgson is a Lecturer in Nuclear Physics in the University of Oxford.

Virus evolution during persistent infection

from C. R. Pringle

IN certain circumstances cells in culture can survive infection by viruses which normally kill them, and the cells continue to multiply still harbouring the infecting viral genome. Nuclear DNA viruses and reverse transcribing RNA viruses persist by integration into the host genome, or by regulated association with the host cell nucleus. The phenotype of the cell may be changed as a consequence and the cell may become oncogenic. However, many cytosolic RNA viruses like the rhabdoviruses and parainfluenza viruses which multiply entirely in the cytoplasm are able to establish a persistent infection in susceptible cells. The molecular mechanisms concerned in the moderation of normally lethal infections in which there is no nuclear involvement are attracting increasing interest. Three factors have been associated with persistent infection. These are temperature-sensitive (*ts*) mutation which impairs virus replication at normal incubation temperature, defective interfering (DI) particles which have portions of the viral genome deleted and suppress the multiplication of homologous complete virus, and production of interferon. The most recent analyses indicate that these factors are interdependent, and that interferon production may hold the balance. Sekellick and Marcus (*Virology* **85**, 175; 1978; **95**, 36; 1979) have shown that many *ts* mutants of the rhabdovirus vesicular stomatitis virus (VSV) are good inducers of interferon in competent cells at restrictive and semi-permissive temperatures, and that the ability of DI particles to protect cells from the lethal action of infectious virus (as distinct from their ability to inhibit viral replication), is mediated by interferon. Interferon levels in persistently infected cultures are usually low, but addition of anti-interferon serum frequently results in an increase in virus production and cell destruction (Youngner *et al.* *J. Virol.* **28**, 6 1978; Ramseur and Friedman *Virology* **85**, 253; 1978). However, persistent infection can be established in cell lines such as Vero and BHK-21 which are genetically deficient in some component of the interferon system. Another important factor may be the progressive selection of a population of semi-refractory cells, because a change in karyotype has been observed in a culture of BS-C-1 cells persistently infected with respiratory syncytial virus (Pringle *et al.* *J. Virol.* **28**, 199; 1978).

Persistent infection has been given new significance by the results of surveillance over five years of a BHK-21 cell culture persistently infected with the Indiana serotype of VSV, with oligonucleotide fingerprinting used to detect changes in the genetic constitution of the virus (Holland *et al.* *Cell* **16**, 495; 1979). The data reveal that the virus has undergone extensive and progressive non-lethal mutational change during this period. The implication is that the mechanisms which moderate virulence to produce persistent infection may simultaneously promote rapid virus evolution.

VSV seems to be genetically stable during repeated cycles of lytic infection. Clewley *et al.* (*J. Virol.* **23**, 152; 1977) found that strains of VSV derived from different field isolates were readily distinguishable by oligonucleotide fingerprinting, although most of the oligonucleotides were conserved. On the other hand one strain maintained independently in six laboratories for 16 yr was unaltered. Likewise Holland and his colleagues could detect no changes in their strain of VSV after eight sequential cycles of cloning in BHK-21 cells, forty cycles of lytic infection in the same cells, and five lethal infections of adult mice. The virus also remained unaltered after growth for 2 months in *Aedes albopictus* cells, although passage in insect cells was known to generate high frequencies of *ts* mutants. These results are not unexpected, however, because 85–90% of all sequence changes would not be detectable by oligonucleotide mapping, because only the larger oligonucleotides (10–15% of the total) obtained by T1 ribonuclease digestion would produce unique spots after polyacrylamide gel electrophoresis in two dimensions.

A very different pattern was observed in the persistently infected culture. This carrier culture had been established by a maturation-defective *ts* mutant of VSV (*ts* G31) in the presence of a DI particle derived from this same mutant. Thus all the genome alterations observed in this culture were a consequence of mutation in the progeny of a single virus particle. The DI particle associated with *ts* G31, which is the smallest of the VSV DIs with approximately 90 % of the genome deleted, was considered to be an essential factor in establishing persistence, although during propagation of the culture it was subsequently replaced by a succession of other DI particles. These DI particles had different oligonucleotide patterns from the original DI of *ts* G31. Infectious virus recovered from the culture after continuous propagation for 14, 21 and 42 months showed 5, 8 and 11 map changes, respectively. At any particular time one DI particle predominated and different clones of infectious virus isolated from the culture at the same passage level had identical oligonucleotide patterns. Because 90 % of sequence changes would have occurred in

the non-unique small oligonucleotides, it is clear that at least 50–100 mutations had occurred by 3.5yr, although only a proportion would result in changes in the amino acid composition of the viral proteins. Nevertheless, it was evident that the VSV genome had experienced extensive and progressive mutational change during persistent infection. The genome changes induced by persistent infection seemed to be stable because virus isolated from the carrier culture after 5yr showed no alteration during ten cycles of lytic infection in susceptible cells.

These observations suggest that persistent infections are a major source of genetic variation for viruses like the parainfluenza viruses and rhabdoviruses, which are unable to acquire new variation by recombination. Holland *et al.* also suggest that persistent infection is an important factor in the evolution of influenza A virus. A given strain may show little variation during periods of acute infection, and extensive sequence alterations (with or without reassortment of subunits) may only accumulate in the interval between pandemics when the virus is sequestered in foci of persistent infection in individual hosts.

One of the most interesting properties of these carrier cells was that they could be propagated in nude (athymic) mice, and it was observed that persistent infection by VSV suppressed the normal tumorigenicity of BHK-21 cells (Reid *et al.* *J. gen. Virol.* **42**, 609; 1979). Analysis of this anti-tumour response may shed light on the difficulties encountered in propagation of certain types of human and other tumours in nude mice. □



A hundred years ago

Intellect in Brutes

The following instance of sagacity in a cat has just been related to me by a friend who knew both the cat and its owner well. The latter, who lived at Ragusa Vecchia, in Dalmatia, was too poor to be able to provide food for the cat; the animal was therefore obliged to cater for himself, and was well known as a thief in the neighbourhood. One day one of the children was being sent off to school without any breakfast; the cat, hearing him sobbing for hunger, immediately went off, and returned with a piece of bread he had stolen from a baker hard by, and brought it to the child. The same thing happened another day, and he came back, dragging along a piece of meat bigger than himself. On crossing the threshold a bit of bone caught in a hole, so puss miaowed till some one came to his help. This same cat, who was constantly catching birds on the roof, slept with some pet birds in a cage without attempting to touch them.

Ragusa, Austria, June 18

MARGARET EVANS

From *Nature* **20**, 3 July, 220; 1879.

C. R. Pringle works at the MRC Virology Unit, Institute of Virology, Glasgow

review article

A unifying model for the G1 period in prokaryotes and eukaryotes

Stephen Cooper

Department of Microbiology, The University of Michigan School of Medicine, Ann Arbor, Michigan 48109

A model to explain the cell division cycle in both prokaryotes and eukaryotes is presented. No specific 'G1 functions' take place during the G1 period, which is merely part of a larger period for the preparation of DNA synthesis which began at the previous initiation of DNA synthesis. A G1 period exists merely because the doubling time of the cells is greater than the sum of the S and G2 periods.

THERE was a time, a few years ago, when the division cycles of prokaryotes and eukaryotes could be clearly distinguished. The division cycle of eukaryotes consisted of three periods—G1, S and G2—whereas prokaryotes had only C and D periods (analogous to the S and G2 periods, respectively) but did not clearly exhibit a G1 period. The S and C periods are analogous when defined as the time of DNA synthesis, and the G2 and D periods are analogous when defined as the time between the termination of DNA synthesis and mitosis or cell division. The analogy between the S and G2 periods and the C and D periods was strengthened by the observations that the lengths of time for these periods were quite invariant with large changes in the growth rate of the various cells¹⁻¹⁰. More recently, however, it has been shown that prokaryotes can have a G1 period¹¹ and that there are eukaryotic cells which appear to be devoid of a G1 period^{10,12-15}. Is there then any common, unified basis for understanding eukaryotic and prokaryotic division cycles?

There have been two different ways of analysing the division cycle of cells, one based on the classic results with eukaryotes and one based on the results obtained with prokaryotes. I believe that this schism is unnecessary and that there is a simple and encompassing view which allows both types of division cycle to be discussed in the same terms. My main premise is that there is no physical reality to the G1 period which requires that it have any particular function; instead the G1 period is merely part of a larger period involved in the preparation for DNA synthesis. This leads to a common, unified model for the cellular division cycle.

Bacterial division cycle

The bacterial division cycle is illustrated schematically in Fig. 1. The C and D periods, analogous to the S and G2 periods, are 40 and 20 min, respectively. That is, it takes 40 min to replicate the DNA of the cell, and there is a period of 20 min between termination of replication and cell division. Figure 1 illustrates the relative rates of DNA synthesis in the cells (the rate being proportional to the number of DNA growing points). Note that for cells growing with a doubling time greater than 60 min there

is a G1 period as well as the S and G2 periods analogous to those in the eukaryotic division cycle. Bacterial cells, however, can grow at rates faster than 60 min per doubling. In these cases Fig. 1 illustrates that there is no G1 period. Instead DNA synthesis occurs during cell division, having started in the previous division cycle. Where the doubling time is less than 40 min DNA synthesis will commence before the completion of previously started rounds of DNA replication. These division cycle patterns are not exhibited by eukaryotes.

The particular pattern of DNA synthesis in the division cycle of a prokaryote therefore is determined solely by the frequency with which DNA replication is initiated. Once initiated, the 'clock' regulating the constant C and D periods determines the subsequent cell division. It is currently assumed that the preparations for DNA synthesis are occurring in the time gap between successive initiations of DNA synthesis. But the preparations for DNA synthesis are occurring continuously and not just in the period designated G1, which will appear when the time for the preparation of initiation of DNA synthesis (equivalent to the doubling time of the cells) is greater than the sum of the C(S) and D(G2) periods.

The nature of the factor(s) responsible for the initiation of DNA synthesis in bacterial cells is not known. Although it is quite reasonable to discuss the phenomenon in terms of a 'hypothetical initiator', circumstantial evidence points to the initiator being synthesised in the same manner as, or in parallel with, the synthesis of cell mass. Therefore, either the cell initiates DNA synthesis when a predetermined cell mass is reached or when some specific initiator, which is a constant fraction of cell mass, is present in the cell. In formal terms, DNA synthesis is initiated when the cell contains a given amount of initiator per origin of DNA to be initiated¹⁶.

Eukaryotic division cycle

In contrast to the bacterial division cycle, the eukaryotic cycle is generally seen to fall into three divisions, a G1 period before DNA synthesis, the S phase defined as the period of DNA synthesis, and the G2 period, defined as the period between DNA synthesis and mitosis plus cell division. (In the following

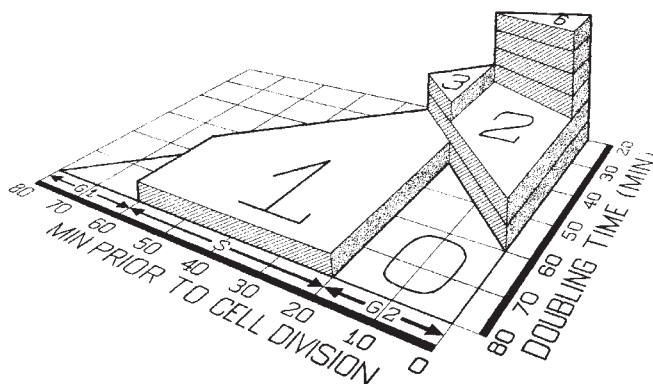


Fig. 1 Theoretical pattern of DNA synthesis during the division cycle of bacteria growing at different growth rates. The vertical axis is the rate of DNA synthesis during the division cycle. Zero DNA synthesis indicates that there is a 'gap' in DNA synthesis and this can be divided into the G1 and G2 periods as indicated. G1 and G2 periods are found when the doubling time is greater than the sum of the S and G2 periods, in this case with doubling time greater than 60 min as the S and G2 periods in *Escherichia coli* B/r are 40 and 20 min, respectively¹. As the graph extends forward to larger and larger doubling times the G2 and S periods remain constant, and the G1 period grows. This portion of the figure is compatible with both prokaryotic and eukaryotic observations. Bacteria can grow with doubling times less than 60 min, however, and the pattern of DNA synthesis during the cycle is altered so that there is no G1 period and the late gap is shortened. The synthesis of DNA is necessarily initiated before cell division, but the time between the termination of a round of DNA replication is unchanged and remains at 20 min. Thus, between 60 and 40 min doubling times the initiation of DNA synthesis occurs during the period between termination of DNA synthesis and cell division (that is, in the bacterial G2(D) period). At doubling times of less than 40 min DNA synthesis is continuous with no 'gaps' in synthesis, and multiply-forked DNA molecules appear. Eukaryotic cells usually have doubling times which exhibit a G1 and a G2 (as in the cultures with doubling times greater than 60 min in this figure).

discussion it is assumed that there is a very small and negligible mitosis and cell division period, so the cycle is divided into only three parts.) To summarise a large amount of data, it seems that the S and G2 periods are fairly constant time periods and therefore are excellent analogues of the C and D periods of bacteria¹⁰.

A generally current view of eukaryotic cells is that they, in contrast to prokaryotic cells, have a defined G1 period in which occur events preparatory for DNA synthesis^{14,17,18}. This view is difficult to reconcile with reports that there are eukaryotic cells which are devoid of a G1 period ("G1-less cells")^{14,15}. How can one have cells which have the G1 events but which do not have G1? I will answer this by presenting a simple description of the division cycle, of both prokaryotic and eukaryotic cells, which suggests that the eukaryotic G1 period has no functional significance and can disappear if the cell grows so fast that the S and G2 periods occupy the entire division cycle. In this model, the 'G1 period' is merely part of a larger period of preparation for the initiation of DNA synthesis.

Description of the general model

Consider three reactions which are observed to occur at various times in the G1 period of cells growing at such a growth rate that they have a substantial G1 period. In Fig. 2 they are indicated as a, b and c. The current (eukaryotic) assumption, that G1 is a period of preparation for S, is illustrated in Fig. 2A. If these cells are now grown at a rate where G1 grows appreciably smaller, S and G2 being relatively constant, then the rate of sequential appearance of the various G1 functions will increase rapidly as indicated in Fig. 2B. In the limiting case where G1 disappears, this model produces a paradox in that there is no G1 period during which the G1 function can take place.

The alternative model, proposed here, is illustrated in Fig. 2C. Here the functions a, b and c occur at precisely the same time as in Fig. 2A with regard to the observed G1 period but these functions are occurring as part of a longer preparation period which stretches from the initiation of S to the next initiation of S. There is no experimental or operational distinction between Fig. 2A and 2C. However, if the cells now grow at a faster rate so that G1 is appreciably shortened, then as indicated in Fig. 2D, the three functions a, b and c can occur earlier than G1—even occurring in S or G2 as shown in Fig. 2D. Where the doubling time is so short that there is no G1, the 'G1 functions' would all take place in either the S or G2 periods.

Illustrative applications of the model

Prescott¹⁰ has summarised the large amount of experimental data on the timing of the G1, S and G2 periods in eukaryotic cells. He has even noted that "These studies lead to the conclusion that the initiation of DNA replication is tied to the attainment by a cell of a crucial mass". He has also noted how this is similar to the bacterial model briefly described above (Fig.

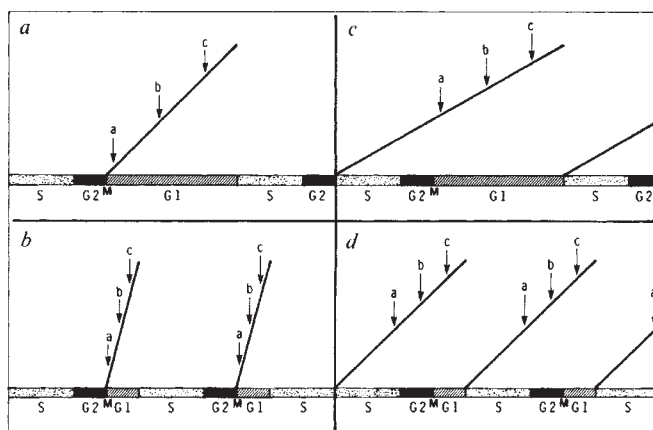


Fig. 2 Comparison of the eukaryotic view of the preparation for DNA synthesis (A, B) and the prokaryotic view of the preparations for DNA synthesis (C, D). A and C are two different interpretations of the observed preparations (a, b, c) for DNA synthesis as experimentally measured during G1. In A they are all part of a period of preparation for S which takes place in G1. In C they are part of a larger period of preparation for S which takes place from one initiation to the next, but with a, b and c only fortuitously occurring during the G1 period. B and D show how each of the models visualises the preparatory events when the rate of growth increases and the G1 period decreases in size. Although this figure is drawn as though there are different events occurring during the period of preparation for DNA synthesis, note that there is no firm and compelling evidence that the initiation of DNA synthesis occurs after a sequence of discrete reactions has occurred. The alternative view, that initiation occurs when some unitary factor has accumulated to some crucial level, is also compatible with this figure with the events a, b and c indicating the achievement of a particular level of the 'hypothetical initiator substance'.

1). I do not wish to repeat his analysis, and rather than review the mass of data on the nature, timing and events occurring in the G1 period, I will show with a few examples how the bacterial model serves to rationalise and explain the data on the nature of the G1 period better than the current eukaryotic model.

For example, Tobey *et al.*¹⁹ have described a specific f1 histone phosphorylation event as occurring in G1 and therefore as a G1-specific event. When cells are growing more rapidly, it would be predicted (although it has not been specifically predicted or stated before, presumably because the possibility has not been studied) that the G1 events would merely occur more

rapidly leading to shorter G1 periods as seen in Fig. 2B. In contrast, the bacterial view presumes that the G1 events occurring in G1 are found in G1 only because G1 exists as an observable phenomenon when the doubling time of the cells is greater than the sum of the S and G2 periods. At a faster growth rate the events which may be observed previously to occur in the G1 period may now occur in the G2 or even the S period. This is clearly illustrated in Fig. 2D. The existence of a G1 event does not support or refute the view proposed here, but study of this phosphorylation reaction in cultures growing at different growth rates might lead to an experimental test of the unified model.

A clearer example is the analysis of eukaryotic cells which do not have a G1 period. Liskay and Prescott¹⁵ have studied such a cell, and have obtained variants from the G1-less cell which contain a G1. They observed that the S and G2 periods did not change and that the appearance of a G1 was entirely accounted for by the increase in the doubling time of the variant cells. I suggest that this is an experimental verification of the idea that the G1 period is produced when the doubling time of a culture is greater than the sum of the S and G2 periods. When Liskay and Prescott selected for cells which had a G1 period, they were actually selecting cells which grew slower and which had less frequent initiations of DNA synthesis. The appearance of the G1 period was not due to any change in the genetic control of the G1 period itself, but merely due to the selection of mutants which had a slower growth rate without any concomitant change in the rate of DNA synthesis (S) or the rate of preparation for cell division following DNA synthesis (G2). Fusion experiments between the slow growing variant and the parental G1-less cell revealed that the variant was a 'deficient condition' which is just what one would expect to produce a slow growing cell. Work by Liskay¹⁴ had shown that the G1-less cells, when fused to temperature-sensitive mutants which had defects in their 'G1 functions', were able to complement the mutants. Therefore, the 'G1-less' cells still exhibited the G1 functions. This would normally seem paradoxical unless one subscribes to the model exhibited in Fig. 2C and D which shows that there is no special G1 period and no special G1 function. Assuming that there are functions which occur sequentially in preparation for DNA synthesis, the model proposed in Fig. 2C and D explains this apparent paradox since it is merely an accident of the cellular growth rate whether the function appears in the G1 period or not. In contrast to the model proposed here, Liskay and Prescott¹⁵ have interpreted the G1-less cells as having 'full expression' (constitutive expression?) of the G1 functions. This is an *ad hoc* explanation which says that G1 functions can be expressed at other times. But it retains the notion of G1 functions.

Implications of the model

The model proposed here presents a unified view of the prokaryotic and eukaryotic division cycles. The main thesis is that the G1 period is merely an artificial construct of the observation, in slow growing prokaryotes and in most eukaryotic cells, of a period of cell growth and synthesis before the S period. The model described above proposes that it is more profitable to look at G1 as part of a larger period, from one initiation of DNA synthesis to the next initiation of DNA synthesis, during which the preparations for DNA synthesis are taking place.

How do eukaryotes and prokaryotes differ?—With regard to the division cycle only, the only difference between prokaryotes and eukaryotes seems to be the apparent inability of eukaryotes to synthesise DNA, that is, have an S period, at the time of cell division. This is possible for prokaryotes because the two nucleoids of the bacterial cell are self-contained and can function while the cell divides. I suggest that this is not the case for eukaryotes, where the completion of the S period produces a single nucleus with twice the diploid DNA content and which separates into two nuclei at mitosis and cell division. It is very likely that mitosis, and the associated condensation of DNA into chromosomes, is not compatible with continued DNA synthesis, and for this reason there is never any DNA synthesis during cell division in eukaryotes. If eukaryotic cells could produce two 'interphase' nuclei following the S period (for example, if the nuclei underwent an immediate mitosis without cell division), then it would be possible for DNA synthesis to occur in these nuclei while the cells were undergoing cell division. This would be analogous to bacterial cells growing with a doubling time of less than the sum of the C and D periods (60 min in Fig. 1).

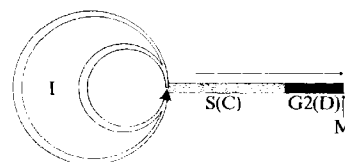


Fig. 3 A schematic representation of the cellular division sequence indicating the period for preparation of initiation, I, which is followed by the S(C) and G2(D) periods finally resulting in M (mitosis). The frequency of initiation is determined by the rate at which preparations for initiation occur, and this is indicated by the two different-sized circles, each representative of a different rate of growth. Note that the preparations for initiation occur continuously and do not start at M.

The cell division sequence—A corollary of my hypothesis is that there is no such thing as the division cycle. The word 'cycle' implies that at cell division something is initiated, the cell cycle. Actually, nothing starts at cell division but it is merely the end of a sequence of events which start with the accumulation of some initiation potential, and which was followed in succession by the initiation of DNA synthesis, the preparations for cell division following termination of DNA synthesis, and the final cell division (Fig. 3). The final cell division is the end of the process and the beginning of nothing. These ideas were expressed in relation to the bacterial cell division a decade ago²⁰.

The ideas described above were developed about a decade ago in discussions with many colleagues. One who deserves special mention is C. E. Helmstetter of Roswell Park Memorial Institute. At that time it was expected that the analysis of eukaryotic cells would eventually follow the logic of the bacterial analysis and therefore an explicit description of a unified model was unnecessary. This has not been the case, hence this brief article. This work was supported by grant DCM 77-14883 from the NSF.

- Helmstetter, C. & Cooper, S. *J. molec. Biol.* **31**, 507–518 (1968).
- Cooper, S. & Helmstetter, C. E. *J. molec. Biol.* **31**, 519–540 (1968).
- Mendelson, M. L., Dohen, F. C. & Moore, H. A. *J. natn. Acad. Cancer Inst.* **25**, 477–484 (1960).
- Sisken, J. E. & Kinoshita, R. *J. biophys. biochem. Cyt.* **9**, 509–518 (1961).
- Young, R. W. *Expl Cell Res.* **26**, 562–567 (1962).
- Defendi, V. & Manson, L. A. *Nature* **198**, 359–361 (1963).
- Cameron, I. L. *J. Cell Biol.* **20**, 185–188 (1964).
- Cameron, I. L. & Greulich, R. C. *J. Cell Biol.* **18**, 31–40 (1963).
- Tobey, R. A., Anderson, E. C. & Petersen, D. F. *J. Cell Biol.* **35**, 53–59 (1967).
- Prescott, D. M. *Reproduction of Eucaryotic Cells* (Academic, New York, 1976).
- Helmstetter, C. E. & Pierucci, O. *J. molec. Biol.* **102**, 477–486 (1976).
- Robbins, E. & Scharff, M. D. *J. Cell Biol.* **34**, 684–686 (1967).

- Liskay, R. M. *J. cell. Physiol.* **84**, 49–56 (1974).
- Liskay, R. M. *Expl Cell Res.* **114**, 69–77 (1978).
- Liskay, R. M. & Prescott, D. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2873–2877 (1978).
- Donachie, W. D., Jones, N. C. & Teather, R. in *Microbial Differentiation* (eds Ashworth, J. M. & Smith, J. E.) 9–44 (Cambridge University Press, 1973).
- Mitchison, J. M. *The Biology of the Cell Cycle* (Cambridge University Press, 1971).
- Pardee, A. B., Dubrow, R., Hamlin, J. L. & Kletzien, R. F. *A. Rev. Biochem.* **47**, 715–750 (1978).
- Tobey, R. A., Gurley, L. R., Hildebrand, C. E., Ratliff, R. L. & Walters, R. A. in *Control of Proliferation in Animal Cells* (eds Clarkson, B. & Baserga, R.) 665–679 (Cold Spring Harbor Laboratory, New York, 1974).
- Helmstetter, C. E., Cooper, S., Pierucci, O. & Revelas, E. *Cold Spring Harb. Symp. quant. Biol.* **33**, 809–822 (1968).

articles

Redshift dependence of colour and space density of radio galaxies

P. Katgert, H. R. de Ruiter* & H. van der Laan

Leiden Observatory, PO Box 9513, 2300 RA Leiden, The Netherlands

Large deep Westerbork samples of radio sources have been optically identified on Palomar Schmidt and Kitt Peak 4-m plates. The radio galaxy population shows no evolution from $z = 0$ to $z \approx 0.25$, then rises steeply in proper density beyond that epoch, attaining enhancement factors ~ 30 at $z \geq 0.5$. These distant radio galaxies show a striking blue excess of as much as two magnitudes.

IDENTIFICATION programs for samples of weak radio sources show that the apparent colours of the faint galaxy identifications (presumably at redshifts of about 0.5) are considerably bluer than expected from redshifting the spectrum of nearby giant ellipticals such as M87. From the present data we cannot decide whether these blue colours are caused by young stellar populations or by (central) non-thermal sources, or possibly both. Using the apparent magnitudes of the identified radio galaxies to derive redshifts we have determined the radio luminosity function (RLF, the space density as a function of radio luminosity) of elliptical galaxies as a function of redshift. For moderate and intermediate radio luminosities, the space density does not increase out to redshifts of 0.2–0.3. However, the space density seems to increase strongly towards higher redshifts, by factors of between 10 and 30 at redshifts of about 0.5. This increase is present even for sources of intermediate luminosities, that is just above the 'break' in the RLF. There is qualitative agreement between RLF estimates using blue and red magnitudes. However, there is a quantitative difference which can be explained if one assumes that the distant radio galaxies are intrinsically bluer than local giant ellipticals, consistent with our empirical colour determinations.

Identifications and colours

The present results are based on optical identification programs of samples of weak sources from the Westerbork Background Survey¹ (BGS) and the 3rd Westerbork Survey² (3rd Wbk), both defined at 1.4 GHz. These surveys comprise more than 100 fields of about 1° diameter each, distributed in an essentially random manner over the northern sky. The majority of the sources in the BGS was identified on III a-J plates taken with the 1.2-m Palomar Schmidt telescope³. When such plates were not available, Palomar Sky Survey prints were used⁴. For about

one-tenth of the sources in the BGS, deep III a-J plates made with the Mayall 4-m telescope could be used⁵. The 3rd Wbk sample was identified on III a-J and 127-04 plates taken with the Palomar Schmidt telescope⁵.

De Ruiter⁶ defined a complete subsample of 785 sources from the BGS. For 155 (that is, 20%) of these sources an optical counterpart was found with $m_b \leq 21.5$ (corrected for galactic absorption). Of the 79 sources for which deep 4-m III a-J plates could be inspected, 14 were identified on these deep plates in addition to those identified on the Palomar Schmidt III a-J plates (a total identification percentage of 47). For 85 of the 238 radio sources in the 3rd Wbk sample (or 36%) an identification was found on the 127-04 and/or III a-J plates. Of the 155 BGS identifications, about 40% are with a stellar object, 45% with a galaxy and 15% with an object for which the type could not be determined. For the 85 identifications in the 3rd Wbk survey these numbers are 20, 55 and 25% respectively.

The criterion for identification was basically one of positional coincidence. From the position uncertainties and the density of objects on the respective plates/prints, we calculate that about 15% of the suggested identifications on Palomar Schmidt plates are spurious. For the identifications on the 4-m plates this fraction is slightly higher, 20–25%. Magnitudes were estimated for the identified objects that were visible on the Palomar Sky Survey prints by measuring image diameters. Magnitude–diameter relationships for galaxies and stellar objects were calibrated using the photoelectric data of Sandage^{7,8}. These calibrations were done independently for the BGS sample (H.R. de R.) and the 3rd Wbk sample (P.K.). Magnitude estimates are probably correct to within 0.5–0.7 mag. As the BGS covers a large range in galactic latitude, a correction for galactic absorption was applied to m_b . This was not done for the 3rd Wbk identifications as these are all at fairly high galactic latitude. For identifications visible only on the Palomar Schmidt III a-J plates a blue magnitude was estimated, taking into account the estimated limit of about 22 mag (corrected for absorption).

Figure 1 shows a colour ($m_b - m_r$) versus magnitude (m_r) diagram for all galaxy identifications visible in the red Palomar Sky Survey prints ($m_r \leq 20$ mag) and for which a blue magnitude was available.

The most striking phenomenon in Fig. 1 is the presence of the very blue objects with $m_r \geq 19$ mag, occurring in both samples. In general, the agreement between the distributions of the BGS and 3rd Wbk samples is very good. We therefore conclude that there are no large systematic differences in the two sets of

*Present address: ESO, Geneva, Switzerland.

magnitude estimates, although we cannot exclude the possibility that both data sets in Fig. 1 contain similar systematic errors. For instance, the very red colours of some galaxies with $m_r \sim 17$ mag are probably due to small (~ 0.5 mag) systematic errors in m_r . However, it is unlikely that the very blue colours of the galaxies with $m_r \geq 19$ mag are the result of systematic errors. Preliminary analysis of area photometry of some objects in the 3rd Wbk sample also indicates that there are no large systematic errors in the magnitude estimates of the faint galaxy identifications. We hope to obtain accurate photoelectrically calibrated photographic photometry for the objects in the 3rd Wbk sample.

Along the upper horizontal scale of Fig. 1 we have indicated the redshifts at which an elliptical galaxy with $M_r = -22.4$ would have the corresponding apparent magnitude (lower scale) in an $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 1$ geometry. The assumption of a constant M_r (independent of redshift) is not unreasonable because the red light is produced by a stellar population which evolves on very long time scales. We used the standard giant elliptical spectrum given by Pence⁹ to calculate the red K -correction needed for this conversion. Figure 2 shows this red K -correction together with the blue K -correction calculated for the same spectrum. Figure 2 also contains the blue K -correction given by Code and Welch¹⁰ on the basis of OAO UV-photometry of M87 (dotted line). These data do not seem to conflict with recent IUE spectral information on M87 (ref. 11).

Assuming an intrinsic colour $M_b - M_r = 1.9$ and using the K -corrections of Fig. 2, one predicts the colour-magnitude relationship shown in Fig. 1. Clearly the faint radio galaxies are very much bluer than expected. We may therefore question whether these objects are elliptical radio galaxies. Their classification as galaxy is reasonably certain because they all have extended (fuzzy) images. In principle, they could be subluminal and hence very nearby objects. However, this does not seem very likely since their angular sizes are what would be expected for galaxies at the redshifts that we have inferred. Therefore they are most likely to be elliptical galaxies, especially if one considers the observed ratio between optical and radio flux.

An interesting question is whether essentially all faint ($m_r = 19$ – 20 mag) radio galaxies are very blue, or rather: how many radio galaxies populate the inaccessible upper right-hand part of Fig. 1. This question cannot really be answered for the BGS sample, because this was defined on III a-J plates so that faint red galaxies are heavily selected against (if they exist). In the 3rd Wbk sample, which should be complete in m_r , there are two

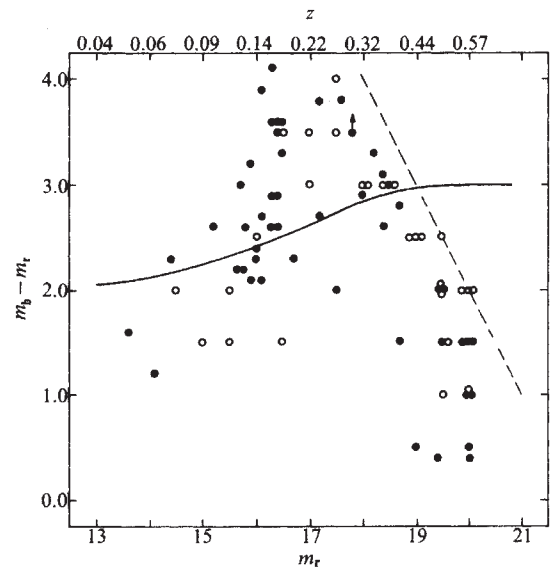


Fig. 1 Colour magnitude diagram of radio galaxies: ●, galaxies from BGS sample; ○, galaxies from 3rd Wbk sample. The solid line is the expected colour-magnitude relationship for a giant elliptical galaxy. The dashed line corresponds to $m_b = 22$ mag, the assumed limit of the III a-J plates

objects (one galaxy and one object of uncertain type) in the range 19–20 mag that are not visible on the III a-J plate. Comparing this with the 12 objects in the same magnitude range below the $m_b = 22$ mag line, leads to the conclusion that most faint elliptical radio galaxies are much bluer than expected.

Our result only applies, of course, to radio-selected elliptical galaxies at redshifts of about 0.5. This 'colour evolution' may occur only in elliptical galaxies that contain a reasonably strong radio source. This would be consistent with the result of Kristian *et al.*¹², who find that the galaxies associated with the radio sources 3C299 and 3C330 (both brightest members of clusters at redshifts of about 0.5) are significantly bluer than the second brightest galaxy in the cluster. Undoubtedly, quite a few of the faint blue galaxies in Fig. 1 are also in clusters but with the present material we cannot tell whether the radio galaxies are bluer than other cluster members which presumably are even fainter. On the other hand, Butcher and Oemler¹³ have obtained photometry of two distant clusters and find that the majority of the galaxies in these clusters are bluer than nearby early-type

Table 1 Magnitude-flux density tables for galaxies

(1) Background Survey identified on IIIa-J 48" Schmidt plates

$\log(S_{1.4}/\text{mJy})$	1.1	1.5	1.9	2.3	2.7
$\Omega_{\text{eff}}(\text{sr})$	7.0×10^{-3}	1.6×10^{-2}	2.2×10^{-2}	2.3×10^{-2}	2.3×10^{-2}
m_b					
17.5–18.5	3(1.2)	1(2.1)	2(2.1)	1(1.0)	0(0.4)
18.5–19.5	2(2.2)	7(3.8)	0(2.8)	1(1.1)	0(0.3)
19.5–20.5	6(3.7)	4(5.3)	2(2.9)	1(1.0)	1(0.3)
20.5–21.5	13(5.1)	11(5.4)	3(2.6)	3(0.8)	1(0.2)

(2) Background Survey identified on IIIa-J 4-m plates

$\log(S_{1.4}/\text{mJy})$	1.1	1.5	1.9	2.3	2.7
$\Omega_{\text{eff}}(\text{sr})$	6.4×10^{-4}	1.2×10^{-3}	1.4×10^{-3}	1.4×10^{-3}	1.4×10^{-3}
m_b					
21.5–23.0	2(0.7)	7(0.5)	3(0.2)	2(0.1)	0(0.0)

(3) 3rd Westerbork Survey identified on 127–04 48" Schmidt plates

$\log(S_{1.4}/\text{mJy})$	1.0	1.4	1.8	2.2	2.6
$\Omega_{\text{eff}}(\text{sr})$	1.3×10^{-3}	4.0×10^{-3}	5.9×10^{-3}	6.2×10^{-3}	6.2×10^{-3}
m_r					
14.25–15.75	2(0.3)	1(0.6)	1(0.7)	0(0.5)	0(0.2)
15.75–17.25	1(0.8)	3(1.9)	1(1.6)	0(0.7)	0(0.2)
17.25–18.75	4(1.8)	2(2.9)	0(1.7)	0(0.6)	0(0.2)
18.75–20.1	1(1.6)	9(2.0)	1(0.9)	3(0.3)	0(0.1)
20.1–21.1	7(1.0)	9(1.0)	4(0.5)	3(0.1)	1(0.0)

The entries in parentheses are predicted galaxy numbers for a bivariate luminosity function identical to the local one for all redshifts.

galaxies. It is not clear whether this indicates that the mix of early- and late-type galaxies in these distant clusters is different from what it is in nearby clusters, or whether the early-type galaxies in the distant clusters are considerably bluer than the nearby ellipticals. Recent work by Kron¹⁴ shows that the colour distribution of faint field galaxies widens strongly towards the blue at the faintest magnitudes. However, the effect seems to occur at least 2–3 mag below the level where we find the blue radio galaxies.

Redshift dependence of the RLF

Before discussing possible implications of these results we consider the other aspect of the identification statistics, which is the information they contain about the radio luminosity function of elliptical galaxies and its redshift dependence. The identification statistics are summarised in the magnitude–flux density tables shown in Table 1. These tables give the number of radio galaxies in various (Δm , $\Delta \log S$) intervals in three samples. In the case of the BGS samples (1) and (2) we used blue magnitudes, while the table for the 3rd Wbk sample (3) is based on red magnitudes.

Note here that Table 1 not only contains genuine galaxy identifications but also a fraction of the identifications with objects of uncertain type. These latter are generally faint identifications, most of which are expected to be galaxies, but which could not be convincingly classified as such from their appearance on the plate. On average, about two thirds of these objects are included in Table 1. In general their inclusion has a relatively small effect although in one or two bins it almost doubles the number of objects.

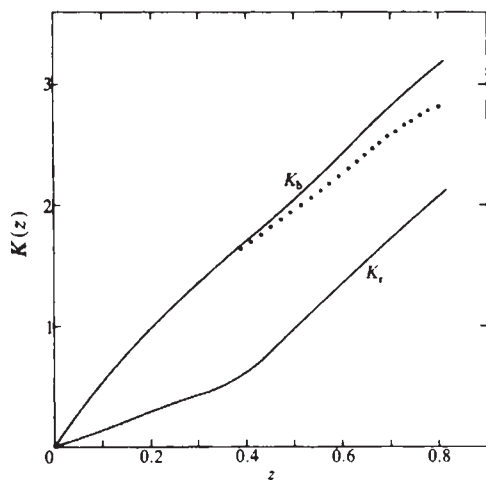


Fig. 2 Red and blue K -corrections based on Pence's standard energy distribution of giant ellipticals (solid lines). The dotted line represents the blue K -correction given by Code and Welch, based on the M87 spectrum.

Table 1 also shows in parentheses the number of radio galaxies expected in each bin if the local bivariate luminosity function of elliptical radio galaxies, as determined by Auriemma *et al.*¹⁵, were valid at all redshifts. This bivariate luminosity function gives the space density of ellipticals as a function of radio luminosity P and absolute optical magnitude M . At any redshift the (Δm , $\Delta \log S$) interval transforms into a (ΔM , $\Delta \log P$)-interval, in which the number of objects per unit volume follows direction from the bivariate luminosity function. Integration over redshift and multiplication of the result by the solid angle covered by the survey in the $\Delta \log S$ interval yields the expected number of galaxies.

Table 1 shows that the observed and predicted numbers are in reasonable agreement for $m_b \leq 20.5$ and $m_r \leq 18.75$ mag. The average absolute magnitudes of the objects in question can be calculated from the local bivariate luminosity function¹⁵ and are $M_b = -20.5$ and $M_r = -22.4$ mag for objects above the 'break' in the RLF. Using these values and the K -corrections of Fig. 2 the above apparent magnitude limits can be converted into redshift limits of ~ 0.3 and ~ 0.4 respectively. Only at fainter magnitudes (at higher redshifts) do the predictions fall short of the observed numbers, by an average factor of, say, 10 to 20.

The observations in Table 1 can be used to estimate the RLF if we ignore the dispersion in absolute magnitude at a given radio luminosity. Magnitude intervals can then be transformed into redshift intervals and, using the solid angles, into volumes, while flux densities transform into luminosities. The result of such a procedure is shown in Fig. 3 (the full curves shown represent our best estimate of the local RLF of elliptical galaxies, based on work by Colla *et al.*¹⁶ and Meier *et al.*¹⁷).

Figure 3b and c give the RLF estimates obtained from samples (1), (2) and (3). The average uncertainty in a single point is shown by the 'error bars'. Estimates based on a single identification are enclosed in parentheses. Different magnitude intervals are indicated by different symbols and nominal average redshifts are given. It is again clear that for redshifts below say 0.3, our estimates do not differ significantly from our best estimate local RLF, which is based on ellipticals with redshifts of ≤ 0.1 . This conclusion extends a similar one by Meier *et al.*¹⁷, who found no evidence for changes of the RLF out to redshifts of about 0.1. By a different, indirect method Robertson¹⁸ also found little evidence for 'evolution' of the RLF below redshifts of ~ 0.2 – 0.3 .

Both samples (1) + (2) and (3) show a strong increase of the space density in the high-redshift intervals. The results of the several samples are compared schematically in Fig. 3d. Although the three samples indicate the same general dependence of the RLF on redshift, sample (3) implies a much less drastic increase than do samples (1) and (2). We consider these differences significant and have investigated several possible explanations.

Colour and density evolution

Note that to first order an error in a zero-point of a magnitude scale displaces RLF estimates in a direction parallel to the local RLF, and hence cannot explain the differences in Fig. 3d. Different alinearities in blue and red magnitude scales could produce an effect like the one observed. For instance, assuming the results of samples (1) and (2) to be correct we find that the limits of the last m_r intervals should be about 19.3 and 20.1 mag, rather than the estimated 20.1 and 21.1 mag. Conversely, the difference in the limits of our III a-J plates from the 48-inch Schmidt and the 4-m telescopes would have to be > 2.5 mag to produce consistency between samples (2) and (3). This is out of the question.

A more plausible explanation is, however, that the relationship between m_b and redshift is not what we assumed it to be on the basis of spectra of nearby ellipticals. In particular, if the UV spectra of the elliptical galaxies are more intense at higher redshifts, we would have underestimated their redshifts. In that case, the objects of sample (2) could be at redshifts of 0.7–0.8 rather than 0.5. This explanation produces consistency between the several RLF determinations and predicts a blue excess of the distant radio galaxies. The observed colour distributions discussed above and shown in Fig. 1 indeed show the faint radio galaxies to be markedly bluer than nearby ellipticals.

Note that the observed colours pertain to *bona fide* galaxy identifications, so that identifications of uncertain type are not represented in Fig. 1. The fact that a very small fraction of these galaxy identifications must be spurious does not change the result. Spurious identifications and identifications of uncertain type enter the determinations of the space density to some

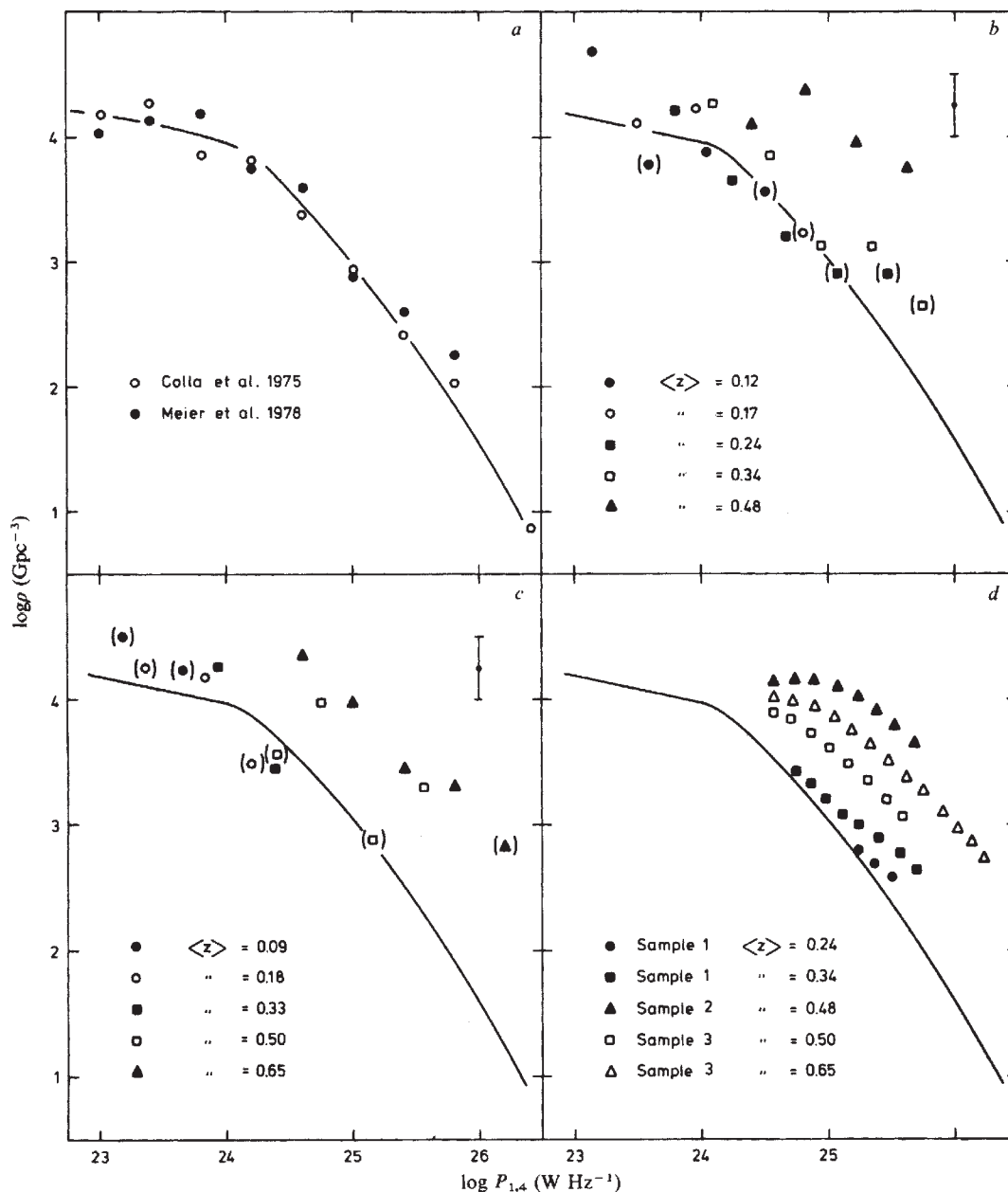


Fig. 3 The radio luminosity function (RLF) of elliptical galaxies: *a*, the local RLF; *b*, RLF estimates from the BGS samples; *c*, RLF estimates from the 3rd WbK sample; *d*, schematic comparison of *b* and *c*. The uncertainty in the individual estimates is shown by the 'error bars'. Nominal average redshifts have been indicated.

extent. The error bars in Fig. 3 include the uncertainties due to these effects, and we therefore think that the differences between the various determinations, which we interpret as evidence for blue colours of the very faint radio galaxies, are significant.

Obviously, the present data should be confirmed and improved. Better magnitudes and colours are required, as well as direct redshift determinations. Nevertheless, the data as they stand indicate that the last 25–30% of the lifetime of the Universe has been relatively quiet. The probability of an elliptical galaxy producing a radio source has decreased drastically. At the same time, the energy output in the UV may have decreased considerably. Of course, the two phenomena may be causally related. If the blue excess has a non-thermal origin, this relationship could be very direct; an active nucleus produces both a blue continuum and a strong radio source. If, on the other hand, the blue excess must be attributed to a very high star-formation rate this relationship would be less direct. In both cases an important question is whether the strong decrease of activity from the $z \sim 0.5$ to the present epoch is caused by evolution of individual galaxies, or is a manifestation of collective processes in clusters.

The Westerbork Synthesis Radio Telescope is operated by the Netherlands Foundation for Radio Astronomy with financial support from the Netherlands Organization for the Advancement of Pure Research (ZWO). H. de R. thanks ZWO for support. We thank Professor Oort for a critical reading of the manuscript.

Received 18 April; accepted 30 May 1979.

- Willis, A. G., Oosterbaan, C. E. & de Ruiter, H. R. *Astr. Astrophys. Suppl.* **25**, 453–505 (1976).
- Katgert, P. *Astr. Astrophys.* **38**, 87–104 (1975).
- de Ruiter, H. R., Willis, A. G. & Arp, H. C. *Astr. Astrophys. Suppl.* **28**, 211–293 (1977).
- Willis, A. G. & de Ruiter, H. R. *Astr. Astrophys. Suppl.* **29**, 103–116 (1977).
- Katgert, P., de Bruyn, A. G. & Willis, A. G. *Astr. Astrophys. Suppl.* **36**, 213–225 (1979).
- de Ruiter, H. R. thesis, Leiden Univ. (1978).
- Sandage, A. *Astrophys. J.* **178**, 25–44 (1972).
- Sandage, A. *Astrophys. J.* **183**, 731–742 (1973).
- Pence, W. *Astrophys. J.* **203**, 39–51 (1976).
- Code, A. D. & Welch, G. A. Wisconsin Astrophysics Preprint No. 43 (1977).
- Boksenberg, A. *et al. Nature* **275**, 404–414 (1978).
- Kristian, J., Sandage, A. & Westphal, J. A. *Astrophys. J.* **221**, 383–394 (1978).
- Butcher, H. & Oemler, A. *Astrophys. J.* **219**, 18–30 (1978).
- Kron, R. G. thesis, Univ. California, Berkeley (1978).
- Auriemma, C. *et al. Astr. Astrophys.* **57**, 41–50 (1977).
- Colla, G. *et al. Astr. Astrophys.* **38**, 209–223 (1975).
- Meier, D. L., Ulrich, M. H., Fanti, R., Gioia, I. & Lari, C. *Astr. Astrophys.* (in the press).
- Robertson, J. G. *Mon. Not. R. astr. Soc.* **182**, 617–628 (1978).

Solar luminosity and the sunspot cycle

R. H. Dicke

Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, New Jersey 08544

Neither sunspots, flares, other solar activity, nor the solar wind, are found to be responsible for the 22-yr solar cycle period found in the [D/H] climate indicator. This period may be due to a periodic variation in solar luminosity induced by a deeply buried magnetic field.

THE discovery of the 11-yr periodicity in the abundance of sunspots by Schwabe¹ in 1843, had initiated the search for solar induced 11-yr periods in meteorological variables of various kinds, and by 1898 many solar-terrestrial relations had been claimed.² But many questionable conclusions had been based on inadequate statistics³ and it was difficult to see how the weak effects expected from solar activity could significantly influence the weather.

Despite these difficulties, it was believed that there might be a connection between solar activity and the weather and the search for a solar-terrestrial relationship apparently led to the establishment of the Laboratory of Tree-Ring Research at Tucson in 1938. But it then took almost 40 yr of hard work to accumulate the tree-ring data which today represent a very substantial statistical base and yield a convincing indication of a drought cycle with a 22-yr period in the western US.⁴ Although this 22-yr cycle had been indicated previously^{5,6}.

The tree-rings show, statistically, a drought cycle with a 22-yr period and not a 11-yr period. Indications of this 22-yr period, and/or the 11-yr period, have been found in several other variables: in the atmospheric pressure at high altitudes⁷, in temperatures, rainfall and other phenomena⁸⁻¹³, but these conclusions have not been uniformly accepted.

The statistically strongest indication of a sharply tuned 22-yr period in a climate indicator seems to be that found by Epstein and Yapp in the [D/H] isotope abundance ratios measured in the cellulose extracted from two bristle cone pine trees which together span 1,000 yr^{14,15}. The period found is in excellent agreement with the double sunspot period¹⁶. The corresponding peak dominates the power spectrum and this peak has substantially the same breadth as that expected from a sinusoidal oscillation of 1,000 yr duration.

The signal strength, its spectral purity and its close agreement with the sunspot period indicate that these data are significant¹⁶. The suggestion that the 10-yr sampling period of [D/H] data could generate a false periodicity could only be true if successive samples were periodically treated differently in the analysis procedure and this seems unlikely. (Owing to the 10 yr averaging, these data could not provide information about the existence of an 11-yr period.) Assuming that the data are significant, the climate is presumably the source of the 22-yr period in the Epstein-Yapp data, but the details are unclear. Unfortunately, much more information would be needed to obtain a description of a 22-yr variation of the climate.

The advantage of the [D/H] ratio as a climate indicator may be the non-local measure of the climate which it provides. The [D/H] ratio in the precipitated water and incorporated into the plant cellulose depends on the mean surface temperature of the Pacific Ocean where the water evaporates and on the air temperatures where the various precipitations occur.

The 11 and 22-yr periodicities are not the only indications of a solar-terrestrial relationship in the weather. Several independent sources of data suggest that there may be a relationship

between mildness of the climate and the average level of solar activity, for example, the sunspot number averaged over the 11-yr cycle. The more sunspots, the milder may be the climate¹⁷⁻²⁰.

Solar activity and the weather

All the obvious physical means by which the solar activity might influence the weather seem to be insignificant, and the presence of the 22-yr period makes the problem doubly difficult. As measured by number and distribution of sunspots the 11-yr period dominates the sunspot cycle. But when the magnetic polarities of the polar regions and the sunspots are included, the period is 22 yr. On symmetry grounds an alteration of the signs of all magnetic fields should not significantly affect energy balance and energy transport, hence the weather. But the polarity of the Earth's magnetic field does not reverse with the solar cycle and magnetic reversals on the Sun could induce a 22-yr period in the weather, but only if the Earth's magnetic field played a significant part (for example, by affecting the solar wind). The solar wind is estimated by Dessler²¹ to carry energy to the Earth at a rate only 6×10^{-7} of the radiation flux from the Sun. This seems too small to be significant.

Other possible physical effects of solar activity seem to be equally weak. At sunspot maximum $\sim 3 \times 10^{-4}$ of the Sun's surface is covered by sunspot umbrae, but a significant change in radiation flux is not implied. First the period is wrong, 11 yr rather than 22 yr, and second, there is not enough heat capacity in the upper part of the convective zone to store the excess heat if the solar radiation is throttled. To obtain an adequate thermal capacity requires a depth of 25,000 km where the pressure ($\sim 5 \times 10^{10}$ dyn cm⁻²) is $\sim 10^5$ times as great as the typical sunspot magnetic pressure. Various effects of solar activity in the upper atmosphere have been reported, for example, on the vorticity^{22,23}.

Roberts says of the difficulty with solar-terrestrial relations in the weather, "It is hard to imagine a plausible scheme to have this tiny tail wag the huge dog"⁶. But does the "tail wag the dog"? It may be, as the various estimates suggest, that the sunspots, faculae, prominences, flares, solar-wind and all the other external manifestations of the solar cycle have a negligible effect on the weather and that the source of the 22-yr period seen in the [D/H] indicator is a solar luminosity variation generated deep inside the Sun. A significant correlation between *A*, the solar luminosity, and *B*, the solar activity, does not imply that *B* causes *A*, or vice-versa. Perhaps *C* causes both *A* and *B*. If so, what could *C* be?

The solar 'chronometer' and the luminosity

It has recently been shown¹⁶ that there is no statistical support for the long-accepted 'eruption-hypothesis' as a description of the sunspot cycle. Under this hypothesis the sunspot cycle undergoes a large random walk in phase, but this random walk is not found statistically. Instead the sunspot cycle seems to be flexibly coupled to some internal 'chronometer', perhaps to a magnetofluid oscillator. The phase of the sunspot cycle correlates strongly with the peak sunspot number. The large phase

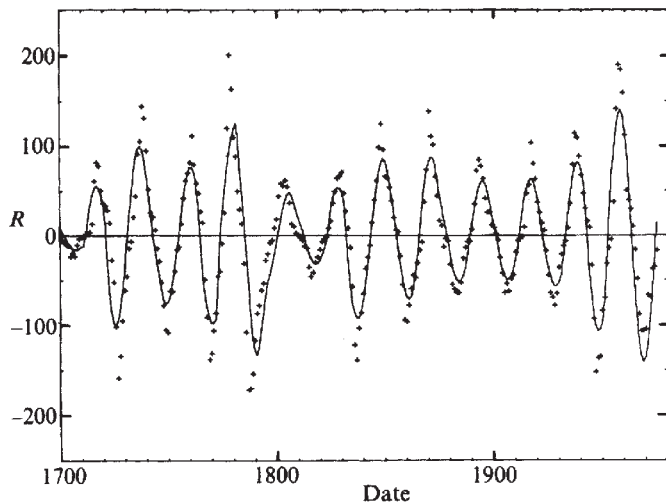


Fig. 1 Sunspot numbers of the (Hale) sunspot cycle and the fitted curve (equation (1)). (See text regarding sunspot numbers.) $\nu = 0.04489 \text{ yr}^{-1}$; $\alpha = 0.015$; $c = 0.8708$.

fluctuation in the cycle may imply a long transit time for the magnetic field responsible for surface activity to float to the surface. If so, the correlation with the peak sunspot number implies that this transit time is less the more active the Sun.

The object C mentioned above might be the 'chronometer'. The Maxwell stress of the magnetic field could affect the energy transport to the surface and modulate the solar luminosity. (A reasonable site for this modulation is near the bottom of the convective zone.) The object C could also be the source of the magnetic field responsible for the sunspot cycle. If so, there are two expectations which might be tested statistically: (i) The solar luminosity should correlate with the sunspot cycle but with a phase delay in the sunspot cycle (the time required for the transport of magnetic flux to the surface). (ii) The correlation of the solar luminosity with the sunspot cycle should be better if the phase fluctuation in the latter is first removed. (The phase shift occurs after the modulation of the solar luminosity by the magnetic field.) These two expectations could be directly tested if one had a measure of the solar luminosity for the past 275 yr.

Solar luminosity, the [D/H] indicator and the sunspot cycle

If the above hypothesis is correct, the 22 yr periodic variation in solar luminosity may be the source of the 22 yr variation in the Epstein-Yapp [D/H] climate indicator¹⁴. This suggests that, as a working hypothesis, we tentatively interpret the indicator as a 'luminosity indicator' and investigate the correlation of the [D/H] data with the sunspot cycle.

As a first step, a curve is fitted to the yearly sunspot numbers^{19,24} for the period 1700–1975 using the previously discussed¹⁶ correlation of the phase of the solar cycle with the strength of the solar activity. According to Eddy¹⁹ the sunspot curve, especially the absolute scale of the curve, is quite unreliable before 1817. Of particular concern is the possible systematic bias of early sunspot numbers through the loss of small sunspots due to poor seeing. From Kiepenheuer's 'visibility function' for sunspots²⁵ an estimated correction factor of ~ 1.3 is obtained. This factor has been arbitrarily applied to all sunspot numbers before 1810. The correction is not secure and all calculations are also carried out with the correction factor omitted.

Figure 1 shows the yearly means of the sunspot number $R(t)$ for the period 1700–1975, with R plotted negative for alternate cycles starting with the first. The curve is a least square fit of

$$H = c\langle R \rangle \cos[2\pi\nu t + \delta_0 + \alpha\langle R \rangle] \quad (1)$$

to the sunspot numbers $R(t)$ where

$$\langle R(t) \rangle = [(R^2(t-5) + \dots + R^2(t+5))/5.5]^{1/2} \quad (2)$$

is an r.m.s. moving mean estimate of the peak sunspot activity. The phase shift $\alpha\langle R \rangle$ represents the previously found correlation of phase with the peak solar activity^{16,26}. The parameters, c , ν , δ_0 and α are adjusted by least squares. The values obtained for ν and α are $0.044886 \pm 0.000054 \text{ yr}^{-1}$, and 0.01153 ± 0.0083 . The formally calculated error estimates are questionable as the residuals show evidence of unmodelled signal.

The double sunspot period p_s , calculated from the above value of ν can be compared with the period previously obtained¹⁶ from the Epstein-Yapp [D/H] data, p_D and with Wittman's²⁷ value, p_w , derived from a combination of telescopic data and a tabulation of naked-eye sunspot observations. These three values,

$$p_s = 22.279 \pm 0.027 \text{ yr}, p_D = 22.360 \pm 0.067 \text{ yr}, \quad (3)$$

$$\text{and } p_w = 22.270 \pm 0.014 \text{ yr}$$

are all in satisfactory agreement. My attempt to obtain a sunspot period from the naked-eye sunspot data floundered in ambiguity¹⁶.

The Epstein-Yapp¹⁴ [D/H] weather indicator, D , is defined as the deviation, measured in parts per thousand, of the [D/H] abundance ratio from a standard oceanic sample. The 1,000 yr interval of time is covered by 100 10 yr samples yielding D_n , centred at t_n , where $n = 1 \dots 100$. These data are filtered to remove low frequency noise by subtracting an approximately gaussian filtered moving mean to give the filtered data

$$D_n^+ = (-D_{n-1} + 2D_n - D_{n+1})/4 \quad (4)$$

D_n^+ is plotted in Fig. 2. The power spectra of both R and D^+ (from Figs 1 and 2) are superimposed in Fig. 3 to show the coincidence of the strong 22-yr peaks.

If the fluctuations of the [D/H] ratio samples were induced by some solar function $f(t)$ lagged in time by τ yr, we would expect the decade time average of $f(t)$ defined by

$$f_n(\tau) = 0.1 \int_{t_n + \tau - 5}^{t_n + \tau + 5} f(t) dt \quad (5)$$

to correlate with D_n and the corresponding $f_n^+(\tau)$ to correlate with D_n^+ . But then the least-square fit

$$c_0 + c_1 f_n^+(\tau) \rightarrow D_n^+ \quad (6)$$

to the N values of D_n^+ could be used to determine c_0 and c_1 for a best fit to D_n^+ . The residuals are

$$\delta D_n^+ = D_n^+ - c_0 - c_1 f_n^+(\tau) \quad (7)$$

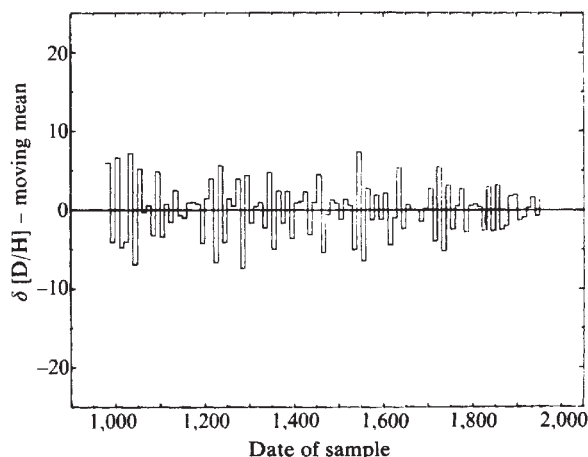


Fig. 2 The Epstein-Yapp Bristle Cone [D/H] data, D^+ (filtered). $D_n^+ = D_n - 1/4(D_{n-1} + 2D_n + D_{n+1})$.

and the unbiased estimate of the residual variance

$$\sigma^2 = \sum_n (\delta D_n^+)^2 / (N-2) \quad (8)$$

provides a measure of the goodness-of-fit, where $(N-2)$ is the number of degrees of freedom remaining after fitting c_0 and c_1 . (The equivalent estimated unfiltered white noise variance is greater than in equation (8) by a factor of $(8/3)$. The error of p_D , equation (3), has been corrected for filtering by the factor $(8/3)^{1/2}$.)

A 22-yr sinusoidal variation in solar luminosity might be expected to lead the 'luminosity indicator' D because of heat storage in the Earth. But from the discussion above, a much longer lag, not a lead, of the sunspot number R relative to D would be expected should the hypothetical internal modulation of the luminosity occur.

Adopting the sunspot number $R(t)$ as the function $f(t)$, equations (4)–(8) yield the unbiased estimate of the residual variance plotted as curve (1) of Fig. 4. The least-square fit equation (6) is calculated for the $N=27$ values of D_n^+ covering the period 1665–1925. This residual variance can be compared with $\sigma_0^2 = 6.275$, the unbiased estimate of the variance of D_n^+ and equivalent to the residual variance obtained when the constraint $c_1 = 0$ is imposed on equation (6).

The above two-least square fits can be considered to represent alternative statistical hypotheses, (1) and (0) and the likelihood ratio of the two fits, with the fit under hypothesis (0) constrained by $c_1 = 0$, is

$$T = (L_0/L_1) = [\sigma^2(N-2)/\sigma_0^2(N-1)]^{N/2} \quad (9)$$

This ratio can be used to test the statistical significance of the various minima of the curves in Fig. 4 using Wilk's method. For large sample, $-2 \ln T$ has a χ^2 distribution with 1 d.f., assuming the truth of hypothesis (0). If a dip of a curve lies below the 1% level the probability of the dip being a statistical fluke (hypothesis (0) being true) is 1% or less. All three dips of curve (1) are significant at the 1% level.

To see if the correlation is improved by eliminating the phase fluctuation from the sunspot cycle, the sunspot number R is replaced by the fitted function H (equation (1)) modified by replacing the phase term $\alpha\langle R \rangle$ by the constant $\alpha\langle \bar{R} \rangle$, where $\langle \bar{R} \rangle = 91.38$ is the mean value of $\langle R \rangle$ averaged over the period 1700–1975. With R replaced by H (modified), curve (2) is obtained. The two dips, occurring with lags of 2.5 and 13.5 yr, are statistically much stronger than those of curve (1).

The Maxwell stress of the magnetic field, not the field strength, should be responsible for the hypothetical modulation of the solar luminosity. To see if an amplitude modulation proportional to $\langle R \rangle^2$ yields an improved fit to D^+ , the H

(modified) used to generate curve (2) is further modified, replacing $\langle R \rangle$ by $\langle R \rangle^2$. The resulting curve (3) has dips at $\tau = 2$ and $\tau = 13$ yr which are slightly deeper than those of curve (2). At $\tau = 13$ yr the likelihood ratio between curves (2) and (3) is 17.2 (it is 2.67×10^3 for curves (1) (3)).

The locations of the minima and their depths, in curve (3), depend on two factors, the phase of the sunspot cycle and the amplitude $\langle R \rangle^2$ in H . To eliminate the phase of the sunspot cycle, f^* in equation (6) is replaced by

$$H_n^+ = [\langle R \rangle_{n-1} + 2\langle R \rangle_n + \langle R \rangle_{n+1}]^2 \cos[2\pi\nu t + \delta] \quad (10)$$

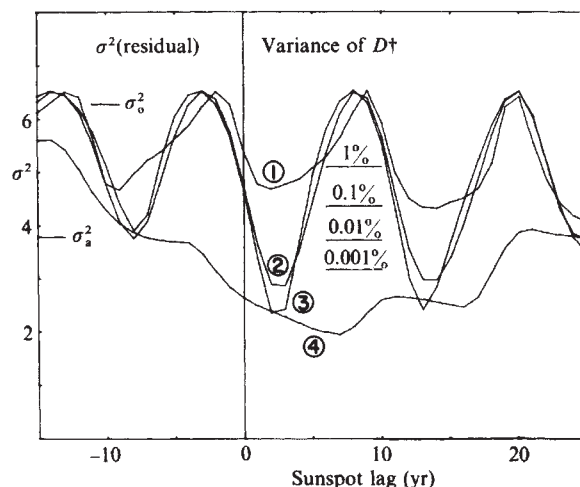


Fig. 4 The residual variance, equation (8), as a function of τ from the fit $c_0 + c_1 f^*$ to D^+ with the following choices of f : ① $R(t)$; ② H (equation (1)) with the phase shift $\alpha\langle R \rangle \rightarrow \alpha\langle \bar{R} \rangle = \text{constant}$; ③, same as ② but with amplitude $\langle R \rangle^2$; ④ f^* from equation (10). The variance of D^+ , σ_0^2 , and the residual variance, σ_a^2 , from equation (11) are also indicated.

with δ adjusted by least-squares. With f^* replaced by H^+ , curve (4) is obtained. The locations of the minima of equations (2) and (3) are little influenced by the amplitude coefficient $\langle R \rangle^2$.

The strongest fit to the data D^+ is derived from equation (10) at the lag $\tau = 7$ yr (see Fig. 5).

Two peripheral questions remain: first, how good a fit to the data D^+ could be obtained by omitting the amplitude coefficient from H (equation (10))? The resulting function

$$f(t) = \cos[2\pi\nu + \delta] \quad (11)$$

gives a residual variance of $\sigma_a^2 = 3.785$, distinctly greater than the minima of curves (2) and (3) and less than the minimum of (1). Second, how do the above results differ if the bias correction of 1.3 is omitted? The sunspot period, to fit equation (1), is $p_s = 22.423 \pm 0.035 \text{ yr}^{-1}$. It has a formal error 30% larger. It agrees with p_D but not with p_w (see equation (3)). Figure 4 is qualitatively unchanged. The dips in curve (1) are only a few per cent higher. The minimum of curve (4) is at $\tau = 6$ yr and is 43% higher. This dip is still significant, at a level of 10^{-5} . The conclusion reached is that eliminating the bias correction makes both the fit of equation (1) to the sunspot data and the fit of equation (6) to the D^+ data somewhat worse, but that, except for the 0.7% disagreement between p_s and p_w (a 4σ discrepancy), the general conclusions reached in this paper are unaffected.

We return now to the expectations (i) and (ii). (i) Curve (1) of Fig. 4 shows a statistically significant correlation of 'luminosity' with R lagged, assuming that D^+ is an indicator for a 22 yr periodicity in the luminosity of the Sun. (ii) Eliminating the variable phase of the sunspot cycle greatly improves the fit to the 'luminosity indicator' D^+ . The phase lags (minimum σ^2) are substantially unaffected but the significance levels of the fits are very much better (0.001%) at the lags of $\tau = 2$ and 13 yr. The

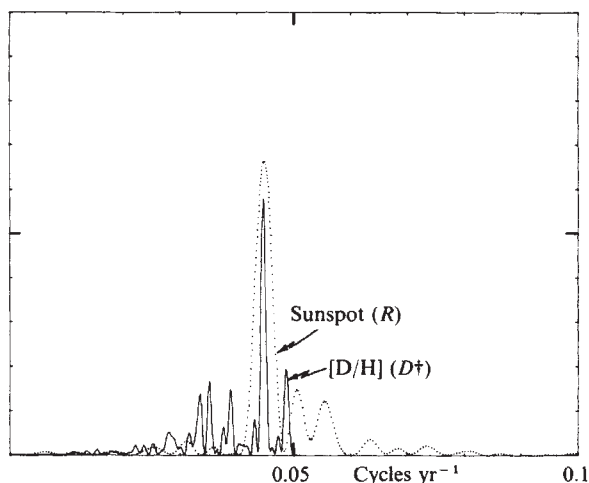


Fig. 3 The power spectrum of R from Fig. 1 (dotted line) and that of D^+ from Fig. 2 (solid line).

inclusion of the amplitude coefficient $\langle R \rangle$ (or $\langle R \rangle^2$) gives a distinctly better fit than is obtained assuming a simple periodic function for $f(t)$.

The phase fluctuation in R is not detectable (at a 2σ level) in D^+ . This is shown by simultaneously fitting H^+ from equation (1) and H^+ modified to eliminate the phase fluctuation to D^+ . The regression coefficients are respectively 0.007 ± 0.029 and 0.032 ± 0.008 . (The errors are corrected by the factor $(8/3)^{1/2}$ for noise filtering.) The first term is not present in D^+ to a measurable extent.

The two lags, $\tau = 2$ and 13 yr of curves (2) and (3) are equally strong statistically but a delay of 2–3 yr, for the eruption of magnetic field at the solar surface, after the modulation of the luminosity is inadequate to permit the large variation in phase seen in the sunspot cycle. This suggests that ~ 13 yr is required for the magnetic field to float to the surface.

Discussion

The 22.36 yr period in the $[D/H]$ climate indicator (D) could be a statistical fluke. But if the 22.36 yr periodicity is meaningful, its close correspondence with the period of the sunspot cycle (Fig. 3), strongly suggests a causal relationship with the solar cycle.

If the solar cycle is driving this periodicity in D , neither sunspots, faculae, flares, other solar activity, nor the solar wind can be responsible. Because all of these surface phenomena exhibit the large phase fluctuation discussed above whereas it has been shown that this phase fluctuation is not present (at a 1σ significance level) in D .

If this 22 yr period in D is meaningful, is due to the Sun, and is not due to any surface or external activity of the Sun, what induces it? A periodicity varying luminosity is a feasible source of variation in the climate, but then the variation must be induced in the deep interior of the Sun, probably towards the bottom of the convective zone and probably by magnetic fields. The work of Epstein and Yapp^{14,15,28} has established that D is sensitive to variations in the climate.

We have shown that the Epstein–Yapp indicator D^+ and the sunspot numbers R are consistent with the following assumptions: (1) that the filtered $[D/H]$ data, D^+ , monitors a 22-yr luminosity variation of the Sun; (2) that the luminosity is modulated by a deeply buried magneto-fluid dynamic oscillator associated with a 'solar chronometer'¹⁶; (3) that some part of this magnetic field floats slowly to the surface where, years later, it is responsible for 'solar activity'; and (4) that the magnetic field arrives at the surface approximately

$$13 - \left(\frac{\alpha}{2\pi\nu} \right) (\langle R \rangle - \overline{\langle R \rangle}) = 13 - 0.041 (\langle R \rangle - 91.4) \text{ yr} \quad (12)$$

after the modulation of the luminosity, where $\langle R \rangle$ is evaluated at the time of arrival at the surface. The hypothetical luminosity modulation is found to have the phase stability of the 'solar chronometer'¹⁶.

The amplitude of the 22-yr period in D (based on all the data and after correcting for the effect of 10-yr averaging) is 3.8 ± 0.5 (see Table 2 of ref. 16). A crudely estimated temperature equivalent of a variation in D ($\delta T/\delta D \approx -0.2^\circ$) is obtained from a comparison of modern values of D and T with those found in North America at the time of the late Wisconsin glacial maximum^{28,29}. Owing to the long time scale of the Wisconsin glaciation and the large magnitude of the variation δD , even the sign of $\delta T/\delta D$ is uncertain when applied to the 22-yr cycle. Nonetheless, the implied times of temperature maximum 1891.0,

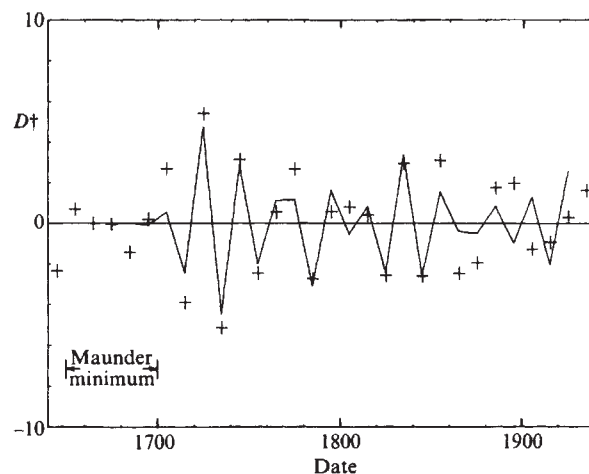


Fig. 5 The fitted function $c_0 + c_1 H^+$, equation (10), and the data points D^+ where c_0 , c_1 and δ are determined for $\tau = 7$ yr. $\nu = 0.04489 \text{ yr}^{-1}$.

1913.3, 1935.7 and 1958.0 (Table 2 of ref. 16) are found to be in rough agreement with the dates of the highplains droughts, 1890, 1913, 1934 and 1954⁶.

For a black-body response of the Earth this estimated temperature amplitude of 0.8°C corresponds to a 1.1% (amplitude) variation of the solar luminosity. More realistic calculations, including albedo feedback effects yield 0.55%, 0.7% and 0.18% for calculations by Budyko, Manabe and Welterald and Schneider and Gal-Chen respectively³⁰.

The only long series of observations of the solar constant, that of Abbot, shows a questionable ~ 23 -yr period³¹. A 0.13% amplitude and a maximum at 1937 are obtained from a fit of a 22.3 yr period to the 1926–46 Abbot data. More recently, balloon flight measurements have given a luminosity increase from 1968 to 1978 of 0.4%, corresponding to an amplitude of the 22-yr period of 0.24%³², assuming the correctness of the D^+ phase.

This work was partly supported by the NSF.

Received 19 March; accepted 14 May 1979.

- Scwabe, H. *Astr. Nachr.* **20** (1843).
- Bigelow, F. H. *U.S. Department of Agriculture, Weather Bureau Bulletin No. 21* (1898).
- Pittcock, A. B. *Rev. Geophys. Space Phys.* **16**, 400 (1978).
- Mitchell, J. M., Jr. Stockton, C. W. & Meko, D. M. *Proc. of Ohio State Univ. Symposium/Workshop on Solar-Terrestrial Influences on Weather and Climate* (in the press).
- Borchert, J. R. *Ann. Ass. Am. Geogr.* **61**, 1 (1971).
- Roberts, W. O. in *Possible Relationships between Solar Activity and Meteorological Phenomena* (eds Banded, W. P. & Maran, S. P.) NASA-SP-366 (NASA, Washington, 1975).
- Ol', A. I. & Sleptsov-Shevelievich, B. A. in *Problems of the Arctic and the Antarctic*, (ed. Treshnikov, A. F.) (National Science Foundation, Washington, 1973).
- Willett, H. C. *J. Atmos. Sci.* **22**, 120 (1965).
- Banded, W. R. & Maran, S. P., *Possible Relationships Between Solar Activity and Meteorological Phenomena*, NASA-SP-366 (NASA, Washington, 1975).
- King, J. W. *Astronaut. Aeronaut.* **13** (1975).
- Siscoe, G. L. *Nature* **276**, 348 (1978).
- La Marche, V. C., Jr *Nature* **276**, 334 (1978).
- Currie, R. G. *J. geophys. Res.* **84**, C2, 753 (1979).
- Epstein, S. & Yapp, C. J. *Earth planet Sci. Lett.* **30**, 252 (1976).

- Epstein, S., Yapp, C. J. & Hall, J. H. *Earth planet. Sci. Lett.* **30**, 241 (1976).
- Dicke, R. H. *Nature* **276**, 676 (1978).
- Bray, J. R. *Nature* **220**, 672 (1968).
- Suess, H. E. *Meteor. Mon.* **8**, 146 (1968); *Science* **171**, 1242 (1971).
- Eddy, J. A. *Science* **192**, 1189 (1976).
- Climatic Change* **1**, 173 (1977).
- Dessler, A. J. in *Possible Relationships between Solar Activity and Meteorological Phenomena* (eds Banded, W. P. & Maran, S. P.) NASA-SP-366 (NASA, Washington, 1975).
- Roberts, W. O. & Olson, R. H. *J. Atmos. Sci.* **30**, 135 (1973).
- Wilcox, J. M. in *Possible Relationships between Solar Activity and Meteorological Phenomena* (eds Banded, W. P. & Maran, S. P.) NASA-SP-366 (NASA, Washington, 1975).
- Waldmeier, M. *The Sunspot-Activity in the Years 1610–1960* (Schulthess, Zurich, 1961).
- Kiepenheuer, K. O. *The Sun* (ed. Kuiper, G. R.) (University of Chicago Press, 1953).
- Lomb, N. R. & Andersen, A. P. *Solar Physics* (in the press).
- Wittmann, A. *Astr. Astrophys.* **66**, 93 (1978).
- Yapp, C. J. & Epstein, S. *Earth planet Sci. Lett.* **34**, 333–350 (1976).
- Emiliani, C. *Science* **168**, 182 (1970).
- Shutt, G. J. & Green, J. S. A. *Nature* **276**, 339 (1978).
- Sterne, T. E. & Dieter, N. *Smithsonian Cont. Astrophys.* **3**, 9 (1958).
- Kosters, J. J. & Murracay, D. G. *Geophys. Res. Lett.* (in the press).

$^{143}\text{Nd}/^{144}\text{Nd}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios from the Azores and their significance in LIL-element enriched mantle

C. J. Hawkesworth*, M. J. Norry†, J. C. Roddick & R. Vollmer‡

Department of Earth Sciences, University of Leeds, Leeds, UK

Basalts from Sao Miguel are displaced to higher $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, and have a significantly shallower slope, compared with the main correlation between $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ for most mantle-derived volcanic rocks. Available data on young magmatic rocks indicate that large ionic lithophile (LIL)-element enrichment in the upper mantle involves two readily distinguishable components which reflect different histories.

MOST oceanic volcanic rocks have higher $^{143}\text{Nd}/^{144}\text{Nd}$ and lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratios than the present-day composition for the model bulk Earth¹. This implies that the oceanic upper mantle has been depleted in Rb and light rare earth elements (LREE) for much of its history. However, there are areas in the ocean basins where volcanic rocks are clearly not derived from a LIL-element depleted source. These are characterised by high concentrations of LIL-elements, high Rb/Sr, Rb/K and Ce/Yb ratios relative to most oceanic lavas, and are referred to as mantle 'hot spots' or 'plumes'^{2,3}. $^{143}\text{Nd}/^{144}\text{Nd}$, $^{87}\text{Sr}/^{86}\text{Sr}$, and REE analyses from one such 'hot spot', the Azores (38°N 25°W), are compared with those from other LIL-element enriched areas. Constraints on models for the nature and timing of enrichment-depletion events in the upper mantle are then discussed.

Sample analyses

The samples analysed are all from the island of Sao Miguel (Fig. 1). This is the largest of the islands in the Azores and it was selected because it has been the subject of extremely detailed geological⁴ and petrochemical work (ref. 5 and Marriner, Norry and Gibson, unpublished data). Its lavas are richer in LIL-elements than those from the other islands of the Azores at the same degree of fractionation; and they were also known to exhibit a larger range in $^{87}\text{Sr}/^{86}\text{Sr}$ (0.7035–0.7052) (ref. 6).

Eleven samples (Table 1), all alkali basalts, were chosen from the collections made by Gibson and Walker. They are from the stratovolcanoes of Aqua de Pau and Sete Cidades and the fissure eruptions between them, and are associated with large volumes of trachyte and hawaiite. The lavas used in this study contain more than 5% MgO and have olivine, clinopyroxene, and sometimes plagioclase, phenocrysts. K_2O ranges from 1.5 to 2.6% and like the other incompatible elements, its concentration is not a simple function of fractionation.

$^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ were analysed by techniques reported previously⁷ and range from 0.7032 to 0.7052 and from

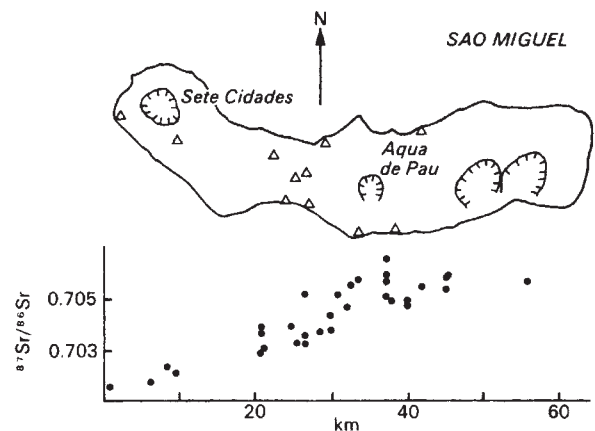


Fig. 1 Sketch map of Sao Miguel in the Azores illustrating the localities of the samples analysed in this study, and the marked increase in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios from west to east across the island. Data from this work, Marriner, Norry and Gibson (unpublished), and ref. 6.

0.51300 to 0.51276 respectively (Table 1). They exhibit an inverse correlation between $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ but have a slope which is much shallower than that observed in lavas from MOR and islands such as Iceland, Bouvet, and Tristan da Cunha⁸ (Fig. 2). They might, therefore, reflect high level contamination⁹ by either seawater or oceanic sediments but for the following reasons we find such an explanation unconvincing.

(1) All the samples were collected from subaerial flows which have not been in contact with seawater. Thus any interaction with seawater (or sediments) presumably takes place within the magma chamber(s).

(2) Assuming that ocean water has $^{87}\text{Sr}/^{86}\text{Sr} = 0.7092$, $^{143}\text{Nd}/^{144}\text{Nd} = 0.5124$ (ref. 10 and Hawkesworth and Elderfield,

Table 1 Alkali basalts sample analyses

Sample*	$\text{K}_2\text{O}^\dagger$	$\text{MgO}^\dagger$	Rb†	Sr†	Rb/Sr	$^{87}\text{Sr}/^{86}\text{Sr}^\ddagger$	Sm/Nd§	$^{143}\text{Nd}/^{144}\text{Nd}^\parallel$
SM 18	1.98	8.54	44	716	0.061	0.70430±4	0.201	0.512886±18
SM 23	1.77	5.57	38	1001	0.038	0.70514±3	0.202	0.512764±20
SM 30	1.93	7.17	43	791	0.054	0.70433±5	0.208	0.512847±16
SM 49	2.65	6.54	59	794	0.074	0.70522±4	0.191	0.512808±18
SM 62	1.43	13.9	37	607	0.061	0.70481±5	0.199	0.512797±30
AZ 1034			48	805	0.060	0.70320±5	0.187	0.513002±20
AZ 1039	1.59	9.17	34	646	0.053	0.70406±5	0.206	0.512910±22
AZ 1322			33	576	0.057	0.70421±3	0.214	0.512872±20
AZ 1324	2.13	5.09	48	906	0.053	0.70391±4	0.202	0.512904±26
AZ 1704			36	612	0.059	0.70348±4	0.200	0.512921±18
AZ 1720	2.51	8.57	55	517	0.106	0.70459±6	0.200	0.512859±20

* All alkali basalts.

† Data from Marriner, Norry and Gibson (unpublished).

‡ Present day $^{87}\text{Sr}/^{86}\text{Sr}$ ratios normalised to E-A, = 0.70800.

§ From Table 2.

|| Present day $^{143}\text{Nd}/^{144}\text{Nd}$ ratios normalised to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$ (BCR-1 = 0.51266±2).

Present addresses: * Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes, UK; † Department of Geology, The University, Leicester, UK; ‡ Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881.

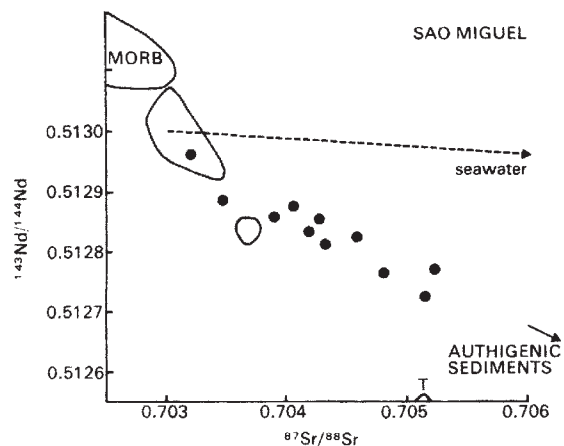


Fig. 2 $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ for basalts from Sao Miguel. I, Iceland; B, Bouvet; T, Tristan da Cunha⁸. BCR $^{143}\text{Nd}/^{144}\text{Nd} = 0.51262$, E-A, $^{87}\text{Sr}/^{86}\text{Sr} = 0.70800$.

unpublished data), and $\text{Sr}/\text{Nd} = 10^6$ (ref. 11), then mixing it¹² with magma of composition 0.705 and 0.5130 results in the almost horizontal trend shown on Fig. 2. This does not seem to be responsible for the slope of the Sao Miguel data.

(3) Given an average of 725 p.p.m. Sr in the basalts and 8 p.p.m. in ocean water, very large water/rock ratios (perhaps 30–50) are implied if the compositions of the basalts are controlled by seawater interaction. Moreover, in contrast to the MOR environment this large volume of water would have to be heated to magmatic temperatures.

(4) Sao Miguel rises some 3,000 m above the floor of the mid-Atlantic Ocean. Moreover, a 1 km drill hole into pillow lavas on the island did not encounter any sediments. Thus for sedimentary material to act as the contaminant it must be dragged down into the magma chamber (presumably along the transform feature), heated to magmatic temperatures, and then mixed in without affecting the primitive nature of the basalts.

It is apparently difficult to attribute the correlation between Nd and Sr isotopes in the Sao Miguel lavas to high level contamination by seawater or oceanic sediments. We believe, therefore, that their isotope geochemistry reflects processes operating within the upper mantle.

There is no simple relationship between the trace element and isotope geochemistry of the lavas studied from Sao Miguel. Eight REE have been determined on each sample using standard isotope dilution techniques¹³ (Table 2). They are all highly enriched in LREEs ($\text{Ce}_N/\text{Yb}_N = 14.2\text{--}10.0$) and only three samples contain detectable Eu anomalies which presumably reflect minor accumulation/fractionation of plagioclase phenocrysts. Considering only the samples without Eu anomalies their Rb/Sr ratios fall in the narrow range of 0.053–0.061 while $^{87}\text{Sr}/^{86}\text{Sr}$ varies from 0.7032 to 0.7048. Similarly $^{143}\text{Nd}/^{144}\text{Nd} = 0.51300\text{--}0.51276$, but Sm/Nd only varies from 0.19 to 0.21. Thus, although we emphasise that there is no clear relationship between isotope composition and trace element ratios, the narrow range of the latter means that they would take 3,000–5,000 Myr to generate the observed range in Nd and Sr isotopes. This is clearly unlikely, and the variation in $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{87}\text{Sr}/^{86}\text{Sr}$ in the Sao Miguel basalts, therefore, probably reflects some recent mixing process.

Further support for this conclusion may be found in the striking regional variation in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios across the island (Fig. 1). It suggests that mixing of magmas or sources took place beneath Sao Miguel and that the high $^{87}\text{Sr}/^{86}\text{Sr}$ component becomes increasingly important in the centre and east of the island. The low $^{87}\text{Sr}/^{86}\text{Sr}$ component has Nd and Sr isotope ratios which lie on the main trend of the Atlantic Ocean volcanic rocks (Fig. 2).

Discussion

Combined Nd and Sr studies have shown that lavas from LIL-element enriched source regions often have isotope compositions which indicate that these sources have actually been depleted in LIL elements for much of their history. Thus when discussing the trace element characteristics of a portion of the upper mantle a clear distinction must be made between those deduced from trace element chemistry in basalts and those inferred from isotope composition.

The isotope results presently available suggest that LIL-element enriched volcanic rocks may usefully be divided into two groups (Fig. 3). Group I consists of the main trend of volcanic rocks on which estimates of the present day $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of the bulk Earth rocks are based¹. It is characterised by a steep slope on $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ and notably, most of

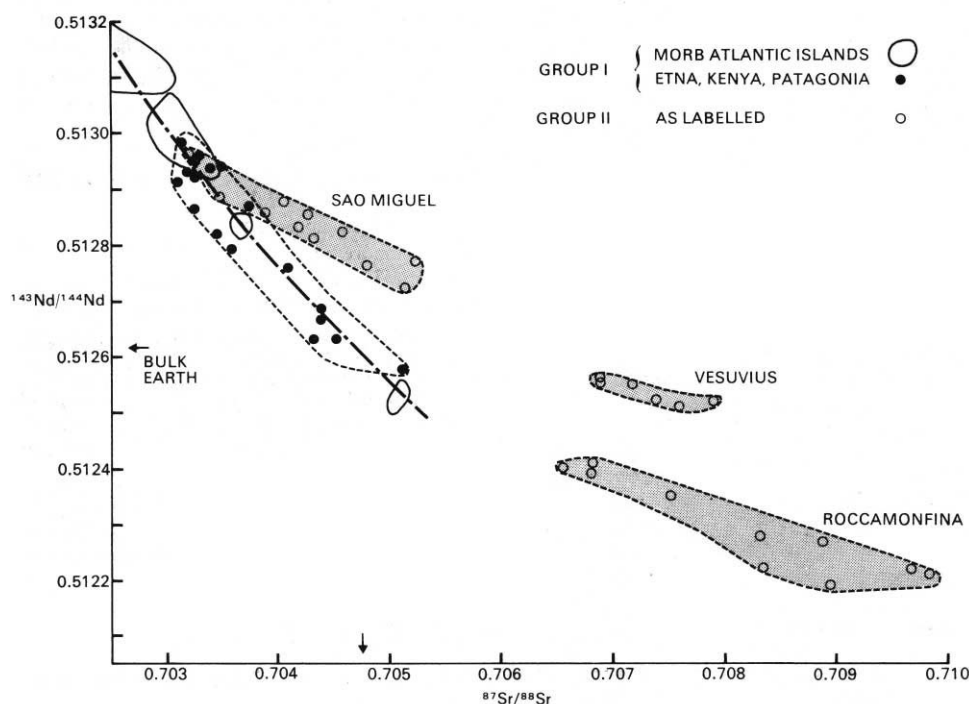


Fig. 3 $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ for basalts from oceanic and continental environments. MORB and Atlantic Islands as in Fig. 2. Results from refs 7, 8, 14, 21, 24 and this work. Present day bulk Earth composition after De Paolo and Wasserburg¹⁵ and O'Nions *et al.*⁸.

the volcanic rocks from both oceanic and continental environments have present Nd and Sr compositions indicative of LIL-element depletion (relative to the model bulk Earth) for long periods of time.

Group II (Fig. 3) is distinguished because its lavas fall on trends of significantly flatter slope which plot to the right of the main trend of oceanic and continental volcanics (group I) on the $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ diagram. Hence they might all reflect contamination; by seawater, sediments, or continental crust. Reasons for rejecting such a hypothesis for Sao Miguel have been outlined above. Hawkesworth and Vollmer¹⁴ argued for Roccamonfina that the positive correlation between $^{87}\text{Sr}/^{86}\text{Sr}$ (0.7064–0.7096) and Sr content (800–2,000 p.p.m.) in its primitive lavas is most unlikely to have been generated, or affected significantly, by interaction with continental crust. There are much less data from Vesuvius¹⁴ but by analogy with Roccamonfina the geochemistry of its lavas are also thought to have been largely unaffected by crustal contamination. Thus we believe that the flat-lying trends of the group II lavas do not reflect contamination but reflect processes active within the upper mantle.

The variation of the trace element ratios relevant to Nd and Sr isotopes can be simply represented on a diagram of Sm/Nd versus Rb/Sr (Fig. 4). Note that the broad curve of the data does not include the point for the Rb/Sr and Sm/Nd ratios of the bulk Earth deduced from Nd and Sr isotope studies^{1,8,15}; as might have been expected if the mantle had a simple history of depletion or enrichment.

Also shown on Fig. 4 are the results of batch melting calculations on garnet lherzolite which have been outlined in more detail elsewhere¹⁶. They illustrate that for amounts of melting exceeding 5–10%, little displacement occurs between melts and their source rocks when plotted on Fig. 4, and that source rocks in quadrant 2 have had a multi-stage history. The simplest model necessitates depletion of a portion of the mantle by extraction of at least 0.5% of liquid followed by remelting to form a second liquid which could mix with other depleted portions of the mantle¹⁶ (Fig. 4).

The Nd and Sr isotope results also suggest that the source regions of the group I lavas have had a complex history. Hawkesworth *et al.*¹⁶ pointed out that the isotope variations on the main trend of the group I lavas could not be reproduced by

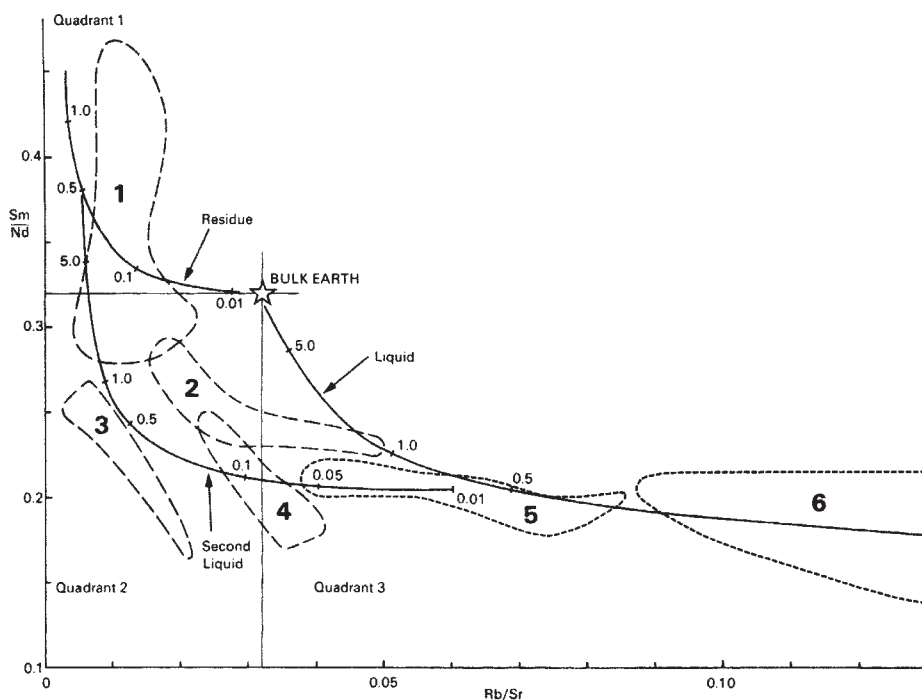


Fig. 4 Plot of Sm/Nd versus Rb/Sr in the more primitive rock types from Iceland (1), Hawaii (2), Etna (3), Patagonia (4), (all group I), and Sao Miguel (5) and Roccamonfina (6) (group II—see Fig. 3). Data from refs 7, 8, 14, 24, 25 and this work. The results of batch melting calculations are from ref. 16 and have the melt fractions in per cent.

Table 2 Rare earth elements

Sample	Ce	Nd	Sm	Eu	Gd	Dy	Er	Yb	Eu/Eu*
SM 18	110	55.2	11.1	3.42	9.47	7.25	3.35	2.59	1.00
SM 23	119	61.8	12.5	4.38	10.8	7.48	3.22	2.30	1.12
SM 30	103	51.5	10.7	3.35	9.31	6.88	3.14	2.35	1.00
SM 49	122	54.4	10.4	3.45	8.73	6.45	2.91	2.18	1.08
SM 62	102	51.2	10.2	3.11	8.74	6.44	2.86	2.25	0.99
AZ 1034	122	57.2	10.7	3.31	8.92	6.35	2.91	2.25	1.01
AZ 1039	89.6	45.5	9.37	2.95	8.14	6.10	2.73	2.03	1.01
AZ 1322	73.7	38.3	8.18	2.59	7.38	—	—	—	1.00
AZ 1324	116	65.0	13.1	4.07	11.3	8.38	3.84	2.95	1.00
AZ 1704	85.9	45.2	9.05	2.88	8.01	5.62	2.58	1.94	1.02
AZ 1720	96.8	45.0	8.98	2.60	7.35	6.12	2.85	2.29	0.95

two-stage models (see Fig. 5) and concluded that some large-scale mixing process was responsible. O'Nions *et al.*¹⁷ suggested that the main Nd and Sr isotope correlation (group I, Fig. 3) results from fractionation of the crust from the mantle throughout geological time, accompanied by continuous remixing of the mantle reservoir. Such a process, while explaining the remarkably good Nd and Sr isotope correlation observed from a large number of areas (group I, Fig. 3) would imply that most mantle materials should have lower Rb/Sr and higher Sm/Nd ratios than the bulk Earth (see quadrant 1, Fig. 4). This is true for MOR lavas and tholeiites from Iceland but many oceanic island and continental basic lavas do not have these characteristics and plot instead in quadrants 2 or 3 of Fig. 4.

Provided that these lavas are produced by amounts of melting exceeding a few per cent, and have not been subject to large amounts of plagioclase fractionation, their source regions will also plot within these quadrants. Prolonged residence in quadrant 2 will result in Nd and Sr isotope compositions plotting to the left of the main isotope correlation (Fig. 3). Lavas with such compositions are rare and when observed in the British Tertiary Province they have been attributed to assimilation of Rb-poor granulite facies crustal material¹⁸. In general, however, most lavas in quadrant 2 of Fig. 4 plot on the main (group I) trend of Nd and Sr isotopes (Fig. 3). Thus we conclude that the enrichment of LIL-elements in their source regions must have taken place comparatively recently so that insufficient time has elapsed to affect their isotope ratios significantly.

Quadrant 3 (Fig. 4) contains lavas derived from source regions with high Rb/Sr and low Sm/Nd ratios; they would develop with sufficient time a flat-lying trend plotting to the right (that is, high $^{87}\text{Sr}/^{86}\text{Sr}$) of the main (group I) trend on the diagram of $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$.

The group II lavas plot in Fig. 4 quadrant 3 and have the flat-lying trend between Nd- and Sr-isotopes that is predicted for source regions which have had variable but high Rb/Sr and low Sm/Nd ratios for a significant period of time. Assuming that trace element ratios in the source region can be estimated from those in the lavas, these would seem to be the ideal situations for recognising 'mantle isochrons'¹⁹ which might 'date' old enrichment events in the upper mantle. However, where such lavas have been analysed for more than one isotope system, the simplest interpretation of the data seems to be that they also reflect geologically recent LIL-element enrichment. At Roccamonfina, for example, the primitive rock types exhibit positive correlations between $^{87}\text{Sr}/^{86}\text{Sr}$ and Rb/Sr, and $^{143}\text{Nd}/^{144}\text{Nd}$ and Sm/Nd, but the calculated ages are 500 and 2,000 Myr respectively¹⁴. Since the $^{207}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ results correspond to an age of almost 0 Myr (ref. 20) we conclude that these 'mantle correlations' are most unlikely to have any real age significance and probably reflect a recent mixing/enrichment event within the upper mantle.

More speculatively it has been suggested that the Nd and Sr isotope compositions from Sao Miguel also reflect some recent mixing process. The high $^{87}\text{Sr}/^{86}\text{Sr}$ component in both examples presumably evolved by virtue of its source having had the high Rb/Sr and low Sm/Nd ratios characteristic of quadrant 3 (Fig. 4) for a significant period of time. It is envisaged that at Sao Miguel this component mixed with material that had LIL-element depleted isotope characteristics similar to Iceland, while at Roccamonfina it mixed with material similar in composition to the model bulk Earth (Fig. 3).

Conclusions

The available Nd and Sr isotope results on recent oceanic and continental volcanics may be conveniently considered in two groups (Fig. 3). Group I includes most of the available oceanic data and is therefore much more significant in terms of volume of mantle sampled. It forms a steep, negative correlation on a graph of $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$, and most of the isotope compositions indicate derivation from source regions which have been depleted in LIL elements relative to the bulk Earth for a considerable time. Recent LIL-element enrichment events are recognised in both oceanic and continental areas, and it is noticeable that (1) the LIL elements have migrated from source regions which also have Nd and Sr isotopes on the trend of the group I samples and (2) that although Rb and Sr abundances are increased during LIL-element enrichment, the Rb/Sr ratios rarely exceed 0.04 (Fig. 4).

Group II samples, in contrast, have shallower slopes and plot to the right of the main correlation (group I) on the graph of $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$. Volumetrically they are clearly much less significant, but an example is now available from both the oceanic and continental environment and their significance

lies in the fact that they clearly involve a different component to those in group I (Fig. 3). We have argued that in the cases studied, this component is most unlikely to be introduced by contamination during magma extraction and that it is therefore present in the upper mantle. It results in high Rb/Sr ratios (>0.04) and correspondingly high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios relative to $^{143}\text{Nd}/^{144}\text{Nd}$. Moreover, since these are often taken to be crustal characteristics, one possible explanation is that this group II component might reflect subducted crustal material which has retained its identity in the upper mantle (O'Nions *et al.* unpublished).

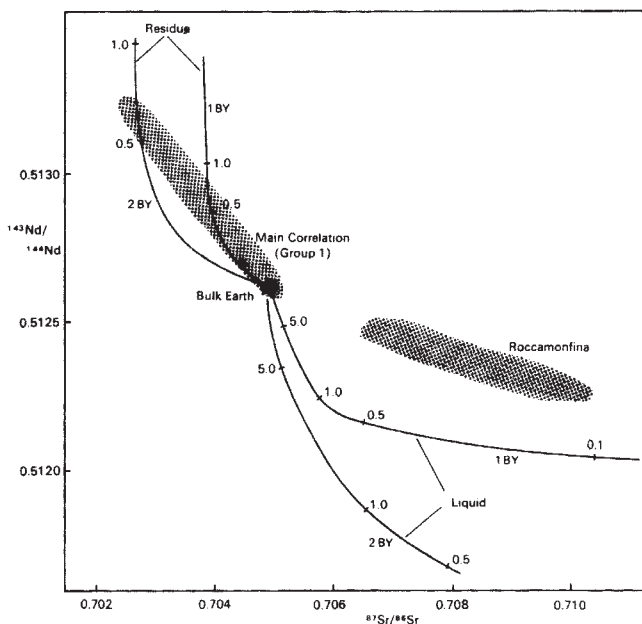


Fig. 5 Nd- and Sr-isotope evolution for two-stage models in which melting occurred 1,000 and 2,000 Myr ago¹⁶. Melt proportions in per cent.

Alternatively mantle metasomatism, as documented from mantle xenoliths in kimberlite by Erlank and coworkers^{22,23}, can result in LIL-element enrichment which will with time generate the relatively high $^{87}\text{Sr}/^{86}\text{Sr}$ and low $^{143}\text{Nd}/^{144}\text{Nd}$ ratios apparently characteristic of the group II end members. It is difficult to determine how long such material needs to be isolated, but note that the data from nodules of metasomatised mantle indicate that $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of 0.705–0.710 can be generated in only 150 Myr (ref. 23).

We thank J. Kramers and R. K. O'Nions for helpful comments. Isotope research at Leeds is supported by grants from NERC and the Royal Society, and we acknowledge NERC Fellowships during this work. We thank M. Greaves, P. Guise, and A. Gledhill for technical assistance.

Received 26 March; accepted 24 May 1979.

- O'Nions, R. K., Evensen, N. M., Hamilton, P. J. & Carter, S. R. *Rev. Earth planet. Sci.* (in the press).
- Schilling, J.-G. *Nature* **242**, 565–571 (1973).
- Schilling, J.-G. *Earth planet. Sci. Lett.* **25**, 103–115 (1975).
- Booth, R., Croasdale, R. & Walker, G. P. L. *Phil. Trans. R. Soc. A* **288**, 271–319 (1978).
- White, W. M., Tapia, M. D. M. & Schilling, J.-G. *Contr. Miner. Petrol.* (in the press).
- White, W. M., Schilling, J.-G. & Hart, S. R. *Nature* **263**, 659–663 (1976).
- Hawkesworth, C. J. *et al. Earth planet. Sci. Lett.* **42**, 45–57 (1979).
- O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13–22 (1977).
- O'Hara, M. J. *Phil. Trans. R. Soc.* (in the press).
- O'Nions, R. K., Carter, S. R., Cohen, R. S., Evensen, N. M. & Hamilton, P. J. *Nature* **273**, 435–438 (1978).
- Högdahl, O. T., Bowen, B. T. & Melson, S. *Adv. Chem. Ser.* **73**, 308–325 (1968).
- Langmuir, C. H., Vocke, R. D., Hanson, G. N. & Hart, S. R. *Earth planet. Sci. Lett.* **37**, 380–392 (1978).

- Hooker, P. J., O'Nions, R. K. & Pankhurst, R. J. *Chem. Geol.* **16**, 189–196 (1975).
- Hawkesworth, C. J. & Vollmer, R. *Contr. Miner. Petrol.* (in the press).
- De Paolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **13**, 249–252 (1976).
- Hawkesworth, C. J., Norry, M. J., Roddick, J. C. & Vollmer, R. V. *USGS Open File Rep.* 78–701, 162–164 (1978).
- O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. *Phil. Trans. R. Soc.* (in the press).
- Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Science* **202**, 743–745 (1978).
- Brooks, C., James, D. E. & Hart, S. R. *Science* **193**, 1086–1094 (1976).
- Vollmer, R. & Hawkesworth, C. J. *Earth planet. Sci. Lett.* (submitted).
- Norry, M. J. *et al. Phil. Trans. R. Soc.* (in the press).
- Erlank, A. J. *EOS* **57**, 597 (1976).
- Erlank, A. J. & Shimizu, N. *2nd Int. Kimberlite Conf.* (abstr.), New Mexico (1977).
- Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Earth planet. Sci. Lett.* **37**, 401–408 (1978).
- O'Nions, R. K., Pankhurst, R. J. & Gronvold, K. J. *Petrol.* **17**, 3–30 (1978).

An exceptional intraplate earthquake sequence in Meløy, Northern Norway

H. Bungum, B. K. Hokland, E. S. Husebye & F. Ringdal

NTNF/NORSAR, PO Box 51, N-2007 Kjeller, Norway

A unique earthquake sequence began in Meløy, Northern Norway, in November 1978, and 10 weeks later about 10,000 shocks had been recorded from a volume not larger than $10 \times 8 \times 6 \text{ km}^3$. The largest earthquake occurred on 28 November, 1978, with an ML magnitude of 3.2 and a maximum MM intensity of VI.

ALTHOUGH Norway is far away from plate boundaries—the nearest ones being the Mohn's and Knipovich Ridges in the Norwegian Sea—the region exhibits a relatively high seismic activity in the context of intraplate earthquake occurrence. Husebye *et al.*¹⁻² have scrutinised all available evidence on the seismicity of this region, including historical records of 400 yr ago. The largest occurring earthquakes were found to have had magnitudes of the order of 6.0–6.5, and broad seismic zones could be delineated, usually in agreement with major tectonic features. In November 1978 a remarkable earthquake sequence began in Northern Norway, where about 10,000 shocks were recorded during the first 10 weeks of instrumental coverage. This unique phenomenon, which is still taking place at a reduced rate is described here.

Background and instrumentation

It is uncertain when the Meløy earthquake sequence started, but the first earthquake positively recorded at NORSAR (the Norwegian Seismic Array, almost 700 km away) occurred on 3 November 1978 and was measured at $ML = 2.4$. ML is the local magnitude calculated from the S_g phase and based on wave attenuation characteristics representative for Scandinavia³. Between 12 and 19 November, ground shakings were often felt, particularly at night time, and some cracks in chimneys and walls were observed. The shocks were accompanied by relatively strong bangs and rumblings. On 16 November NTNF/NORSAR were asked to monitor the area, and two days later the first seismographs (type Sprengnether MEQ-800) were installed. A maximum of five stations were in operation at one time while seven different sites were used (for details, see Fig. 1 and Table 1). Three of these sites—ENG (Enga), NEV (Neverdal) and ORN (Ørnes)—were chosen as permanent locations. Figure 2 shows the microearthquake activity on one of the most active days. Accelerometers, hydrophones and analog and digital tape recording of three-component seismometers were also installed on an experimental basis and operated for short time intervals. (Field work in the rugged Meløy area is very difficult in winter so more extensive monitoring experiments are planned for the summer assuming the earthquake sequence continues until then.) Besides the NORSAR array, several other seismograph installations in Scandinavia have recorded the largest Meløy earthquakes. Whenever possible, the NORSAR recordings have been used to measure the local ML magnitude.

Earthquake sequence in time and space

Minor tremors are not uncommon in the Meløy area, and the first reported rumblings that may be associated with the present sequence date back to around 20 October 1978. While it is

difficult to determine the exact starting date, it seems that the main activity started around 10 November 1978.

The timing of the earthquake sequence between 11 November 1978 and 26 January 1979 is given in Fig. 3 for station NEV and shows the total number of events per 4-hourly time interval, the cumulative number of events (ending at about 7,500), and all events with an ML magnitude of at least 2.0. The main concentration of larger shocks was between 11 and 18 November, and there is also an interesting predominance of night-time shocks during this interval. The activity was high and fairly constant (although with large short-term variations) up to 5 December, then decreased to a minimum around 20 December, whereafter new outbursts of activity occurred on 27–28 December, 3–4, 8–10 and 18–20 January. Comparing the occurrence of the larger events with the total activity we see a repeated pattern of 'main' shocks followed by aftershocks lasting sometimes a few hours, sometimes a few days.

A detailed study of the sequence shows only one example of foreshock activity: the $ML = 2.9$ event on 27 December which was preceded by nine small shocks starting 90 s before the main one. Apart from that event, the main shocks were always preceded by a period of relative quiescence. There are also examples of large sub-sequences not preceded by a main shock,

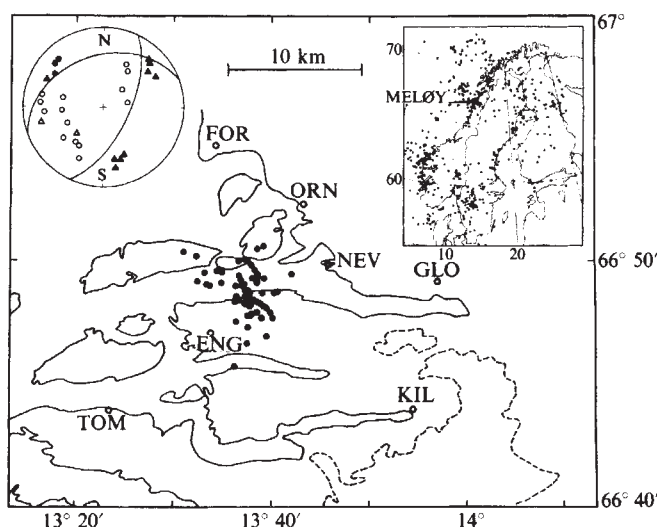


Fig. 1 Map of the Meløy area with the locations of seven seismic stations whose times of operation are given in Table 1. The dotted line to the south-east indicates a glacier (the whole area is very mountainous). ●, Computed epicentres for 66 earthquakes for which from 5 to 10 readings (phases) were available (see text for details) and the corresponding depth range is 3–9 km. The upper right insert shows Fennoscandia with instrumentally located earthquakes for the time period 1951–78, and the location of Meløy is also given. The upper left insert is a composite focal mechanism solution for the Meløy earthquakes, where compressions (solid symbols) and dilatations (open symbols) are plotted in a lower hemisphere stereographic projection. See text for explanation of symbols. The two nodal planes have strikes/dips of $25^\circ/63^\circ$ E and $250^\circ/35^\circ$ N, respectively.

Table 1 The Meløy microearthquake stations with recording time intervals and the respective numbers of recorded events

Station	Code	Start	Stop	Recorded events
Fore	FOR	18.11.78	22.11.78	95
Neverdal	NEV	18.11.78	26.01.79*	7,494
Tømmervik	TOM	19.11.78	23.11.78	34
Kilvik	KIL	21.11.78	27.11.78	175
Enga	ENG	23.11.78	26.01.79*	7,628
Glomfjord	GLO	24.11.78	29.11.78	41
Ørnes	ORN	26.11.78	26.01.79*	2,970

* Station still in operation, March 1979.

The total number of individual events recorded up to 26 January 1979 was 9,954. See Fig. 1 for station locations.

such as on 4 January when 103 microearthquakes were recorded at ENG within an hour with no event magnitude exceeding $ML = 2.0$. The largest number of events recorded in one single day occurred on 2 December with 820 at ENG, 750 at NEV and 270 at ORN.

The hypocentre was located as follows: an initial epicentre estimate was obtained on the basis of S-P time differences using a maximum likelihood technique⁴⁻⁵ and then refining this estimate iteratively using absolute P/S arrival times when such data were available. The associated crustal model was the simplest possible, with a P_s velocity of 6.25 km s^{-1} (P_b or P_n were never observed), and a P/S velocity ratio of 1.78. Depth of focus was obtained by looping over a specified number of depths, choosing the one which resulted in the smallest time r.m.s. value. This procedure has proved to be reliable even with unfavourable conditions such as non-symmetric station configuration, few observations and unreliable station timing corrections.

We have computed hypocentres for 255 events evenly distributed in time throughout the recording period. These were all intermediate magnitude events, as the largest ones could not be read reliably because of large deflections (saturation) on the paper recordings. Between three and five stations and between six and 10 phases (S-P relative time counts as one phase) have been used. Out of 87 hypocentral solutions with five or more phases, 85% have r.m.s. values $<0.20 \text{ s}$ and 45% have r.m.s. values $<0.10 \text{ s}$. For the same events the largest semi-axis of the 95% uncertainty ellipse has a length of $\sim 1 \text{ km}$ (assuming a reading error of 0.1 s on both P and S), and for the 'poorer' other events usually not more than 2 km . In Fig. 1 66 events with at

least five readings (phases), and r.m.s. values $<0.15 \text{ s}$ are plotted. The epicentres are confined within an area of roughly 10 km N/S , 8 km E/W , with a concentration around 66.81°N , 13.63°E . In view of the small location uncertainty the epicentral scatter in Fig. 1 is clearly significant, and is not an artefact of recording or data analysis. The computed depths for the 66 events are in the range $3\text{--}9 \text{ km}$, with 19 within $3\text{--}5 \text{ km}$, 27 within $5\text{--}7 \text{ km}$ and 20 within $7\text{--}9 \text{ km}$. Based on the computations of r.m.s. values against depth for each event, we find that the uncertainty in depth is of the same order as for the epicentral coordinates—usually within $\pm 1 \text{ km}$. There is one important exception here, however, when only stations ORN, NEV and ENG are used, there is poorer resolution between shallow events to the NW and deeper events near the centre of the activated area. Finally, as the epicentres are roughly surrounded by stations, possible small errors in the crustal model are considered unlikely to adversely affect the hypocentre locations.

A study of the hypocentral 'time development' shows that essentially the whole $10 \times 8 \times 6 \text{ km}^3$ volume has been activated for all the time. However, there are indications of some migrations in the earthquake activity, predominantly in a N-S direction. The southernmost and easternmost events have rather shallow depths ($3\text{--}6 \text{ km}$), while events nearer the centre of the area cover the whole depth range of $3\text{--}9 \text{ km}$. The magnitude-frequency relationship for the largest events in the sequence is plotted in Fig. 4, with a non-anomalous slope of 1.13.

We have measured first-motion directions for a large number of events—not an easy task due to a 60 mm min^{-1} time resolution in our seismograms. For KIL, GLO and ORN all readings were unidirectional, while for the four other stations (see Table 1) there was one predominant direction but also observations to the contrary. At this stage it is difficult to decide whether these changes in first-motion directions are caused by either genuine differences in focal mechanisms or in hypocentral locations, or possibly also by reading errors. Notwithstanding these problems we have obtained a composite focal mechanism solution for the Meløy earthquakes as presented in the upper left corner of Fig. 1. The solution is based primarily on four events with locations along a line striking about $N 10^\circ \text{E}$, each one separated by about 2 km . Two of the events were recorded by five stations, one by four stations and one by three stations, while the first-motion directions for the rest of the seven recording sites ($\blacktriangle$) are inferred from other events with nearby locations. The solution indicates normal faulting, where the plane striking $N 25^\circ \text{E}$ and dipping 63°E is the one which gives the best fit to hypocentral solutions, therefore also being the probable faulting plane.

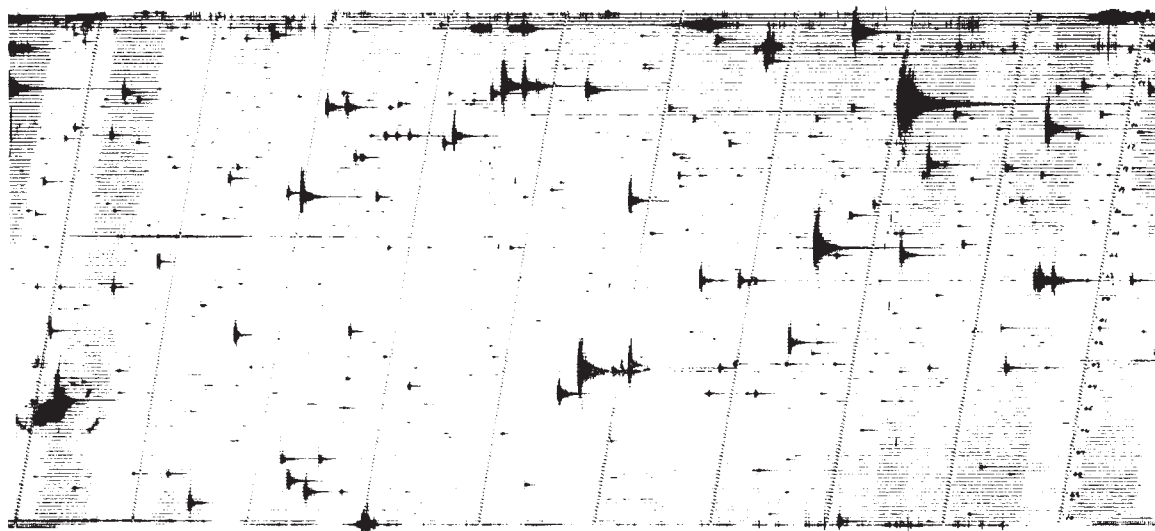


Fig. 2 Reproduction of the seismogram from station ENG covering about 24 h starting at 10.40 GMT on 2 December 1978. The earthquake swarm on the seismogram is an aftershock sequence following a magnitude (ML) 2.7 event at 02.04 GMT the same day, about 540 events were then recorded during the next 9 h until the onset of the reproduced seismogram, from which about 400 new events have been identified. The largest of the earthquakes on the seismogram (at 14.57 GMT) had a magnitude of $ML = 2.1$.

Macroseismic observations

Newspaper ads, instead of conventional macroseismic questionnaires⁷, published during the main earthquake activity in the Meløy area (about 2,000 households along the shore lines in Fig. 1), one on 18 November covering the activity up to then and one on 30 November covering the large events on 28 and 29 November, produced 33 and 63 answers, respectively, from which we have derived a maximum radius of perception of 50 km and a maximum intensity of VI for the largest earthquakes in the sequence. This intensity was assigned to 15 of the 96 answers, all of them <10 km from the focus. In 10 of these cases damages were reported on houses, such as minor cracks in walls and chimneys.

Most of the events above $ML = 2.0$ were felt, and many more were heard. The sounds reported can be classified in three groups: (1) sounds without feeling tremor (on one occasion this was reported as a more or less continuous phenomenon over a period of several hours); (2) sounds associated with tremors, that is, simultaneous with P/S body wave arrivals; and (3) sounds associated with propagation of pressure waves through the air and from the epicentral area. The latter sounds (described as sounding like a load of snow falling off a roof) were quite different from group (2) and for the largest of the earthquakes in the sequence (28 November, $ML = 3.2$) these sounds were also observed/heard (with sufficient timing accuracy) by one of us (B.K.H.).

Discussion

The Meløy earthquake sequence is an outstanding example of intraplate seismic activity⁸. Earthquake swarm activity is not an exceptional geophysical phenomenon, as this type of process is frequently observed in volcanic areas⁹⁻¹⁰ and also at interplate boundaries like the San Andreas fault¹¹. We have not, however, been able to find a clear published counterpart to the Meløy sequence, though there are some similarities to earthquake activities reported in the Mississippi valley¹², Caucasus¹³ and the Garm¹⁴ areas.

The seismicity map for Fennoscandia in Fig. 1 shows that, although the area is not very active on the global scale, it is still far from being aseismic. Moreover, Meløy is located in the middle of a distinct seismicity zone along the coastal area between 65° and 70° N. Within this zone one of the largest events in Fennoscandia so far took place in 1819, the so-called

Lurøy earthquake, with a presumed location of about 50 km south-west of Meløy. Its magnitude is estimated to 6.0–6.5 m_b units, and was followed by an extensive aftershock sequence. Historical records, however, give no indications of foreshock activity.

Further west of the Meløy area (and south-west of the Lofoten Islands) there is another seismicity zone² which incidentally coincides with a passive continental margin (escarpment) dating back to the opening of the Norwegian Sea some 58 Myr ago¹⁵. The Lofoten area where the oldest rock ages are of the order of 3,500 Myr are generally considered a larger Baltic Shield fragment left mainly undeformed during the Caledonian orogeny¹⁶, resulting in the closing of a proto-Atlantic ocean some 400 Myr BP. The seas between the two seismicity zones discussed above are part of an epicontinental basin with maximum sedimentary thicknesses of the order of 8–9 km (ref. 17). In the northern Vestfjorden sub-basin seismic profiling surveys have given clear evidence of block faulting¹⁸. Sometimes a causal relationship has been suggested between the Norwegian Sea opening and the contemporaneous early Eocene uplift of western Fennoscandia which amounted to a maximum of 2 km. This uplift was strongest in the central parts of Norway, tapering off rapidly westwards and gently eastwards.

A spectacular neotectonic feature is found some 200 km to the north-east in Swedish Lapland, namely, the so-called post-glacial Pärve fault¹⁹. The length of this fault, which is 'facing' westwards, is approximately 150 km and the maximum height is 25–30 m. Another neotectonic feature is the presumed glacial rebound of Fennoscandia, where a controversial suggestion is that this rebound has a significant tectonic component²⁰.

The main problem is how to fit the Meløy intraplate earthquake sequence into the tectonic framework of the area²¹. A strike direction of 20°–30° (as inferred from the epicentre distribution as well as the focal mechanism solution) is coincident with both the Caledonian folding axis, the sedimentary basin axes and that of the Pärve fault. Second, the normal faulting with an eastward dipping angle of 63° seems to rule out a causal connection both to the Eocene uplift and the present glacial rebound, having positive gradients for onshore areas. Note that the present uplift with its centre in the Bothnian Bay does not seem to correlate with Fennoscandian seismicity as discussed recently by Husebye *et al.*²; nor do we see a direct relationship to the off-coast sedimentary basin block tectonics

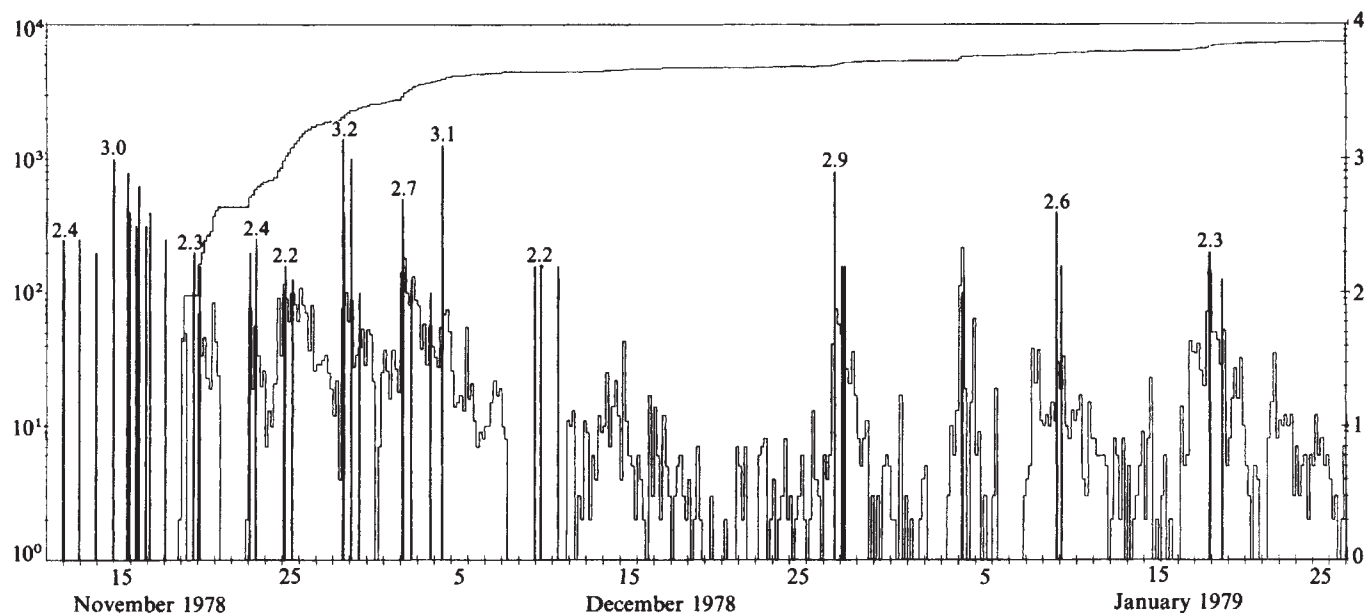


Fig. 3 Time development of the Meløy earthquake sequence as recorded at station NEV (see Fig. 1). The histogram shows the number of events on a 4-hourly basis (non-operational gaps on 21–22 November, 8–11 December 1978 and 2–3 January 1979), including the cumulative number of events and with reference to the left side scale. All events with an ML magnitude at 2.0 or above are plotted separately as vertical lines as referenced to the scale on the right side. For some of these larger earthquakes the magnitude is also given directly.

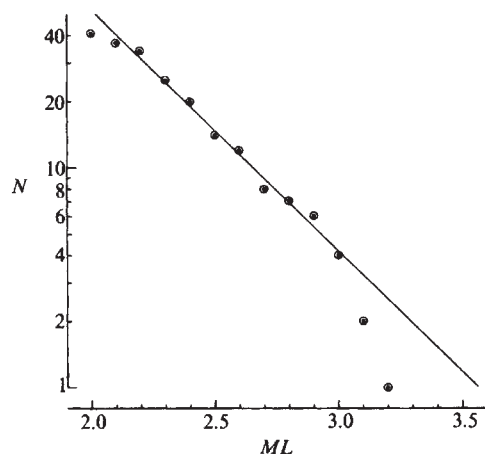


Fig. 4 Cumulative magnitude-frequency distribution for the largest events of the Meløy sequence (plotted individually in Fig. 3). The straight line is fitted by eye and satisfies the Gutenberg-Richter⁶ recurrence formula ($\log N = a - b \cdot ML$) with a slope $b = 1.13$.

where the fault walls nearest to the coast face westward. However, a more subtle relationship to the off-coast sedimentary basins is possible if recent hypotheses on their formation are approximately valid²². The arguments here are as follows: after an initial rifting/graben formation process the sedimentary depositional potential of the area in question is maintained due to continuous subsidence of the area as part of an isostatic compensation mechanism on a lithosphere whose loading response is either elastic or visco-elastic. Such a process could be associated with an 'overshoot' effect on the flanks, and in our particular case normal faulting facing eastwards. Another very generalised explanation is that in Meløy we have an exceptional example of release of remnant stresses associated with the Caledonian orogeny.

One active seismological research area is that of reanalysis of large earthquakes to investigate whether there was evidence of a precursory phenomena. In the case of Meløy the problems are somewhat different, as earthquake activity of the type reported here intuitively will be suspected as being precursory for a significantly larger event. However, in view of the exceptional nature of this intraplate earthquake sequence, it is difficult to argue convincingly either for or against such a hypothesis. If the magnitude-frequency relationship (see Fig. 4) should be used in a predictive context, we would expect a few somewhat larger earthquakes ($ML \sim 3.4-3.6$) to occur in the Meløy area. Also

most processes in nature are multiplicative (exhibit a log normal distribution) so we would expect that a few events should be significantly larger than others.

We know from historical records that the general Meløy area has the potential of large earthquakes². We have no firm evidence, however, of whether or not the present microearthquake activity is precursory to somewhat larger earthquake(s) to take place there in the near future. The large Lurøy earthquake of 1819 is not considered conclusive in this context as the lack of foreshocks in this case may simply reflect poor reporting practices at that time. The Meløy earthquake sequence is not only very interesting but is also puzzling, and hopefully the planned close monitoring of the area in the time to come will shed some further light on the still remaining problems.

We thank the Norwegian Polar Research Institute, Oslo, the Seismological Department, Geodetic Institute, Copenhagen, the Seismological Observatory, University of Bergen, the Norwegian Geotechnical Institute, Oslo, and Det Norske Veritas, Oslo, for cooperation. A Nilsen, NTNF/NORSAR, supervised the initial installation of recording equipment with assistance from the Meløy municipality representatives. We thank O. Fuhr, S. Gregersen, J. Hjelme, K. Hove, Y. Kristoffersen, B. Lillevik, I. B. Ramberg, P. Selnes, T. Søyland-Hansen, O. Strand and S. Vaage for assistance and discussions. This research was supported by the Royal Norwegian Council for Scientific and Industrial Research (NTNF).

Received 9 March; accepted 16 May 1979.

- Husebye, E. S., Gjølsetdal, H., Bungum, H. & Eldholm, O. *Tectonophysics* **26**, 55-70 (1975).
- Husebye, E. S., Bungum, H., Fyen, J. & Gjølsetdal, H. *Norsk. Geol. Tidsskr.* **58**, 51-68 (1978).
- Båth, M., Kulhanek, O., Van Eck, T. & Wahlström, R. *Rep. No. 5-76* (Seismological Institute, Uppsala, 1976).
- Bungum, H. & Kristoffersen, Y. *Nor. Polarinst. Årbok* **1977**, 237-246 (1978).
- Gjølsetdal, H. *NTNF/NORSAR Sci. Rep. No. 2-77/78* (1978).
- Gutenberg, B. & Richter, C. F. *Seismicity of the Earth* (Princeton University Press, 1954).
- Ringdal, F., Husebye, E. S., Sandvin, O. A. & Christofferson, A. *Tectonophysics* **49**, 161-170 (1978).
- Sykes, L. R. *Rev. Geophys. Space Phys.* **16**, 621-688 (1978).
- Klein, F. W., Einarsson, P. & Wyss, M. *J. geophys. Res.* **82**, 865-888 (1977).
- Rikitake, T. *Earthquake Prediction*, 104-107 (Elsevier, Amsterdam, 1976).
- Wesson, R. L. & Ellsworth, W. L. *J. geophys. Res.* **78**, 8527-8546 (1973).
- Stauder, W., Kramer, M., Fischer, G., Schaefer, S. & Morrissey, S. T. *Bull. seism. Soc. Am.* **66**, 1953-1964 (1976).
- Dubruchenko, A. V., Zarskiy, M. N., Vanducheva, N. V. & Shebalin, N. V. in *Earthquakes in the USSR in 1971* (in Russian) 36-45 (Nauka, Moscow, 1975).
- Rautian, T. G., Khalaturin, V. I., Martynov, V. G. & Molnar, P. *Bull. seism. Soc. Am.* **68**, 949-971 (1978).
- Talwani, M. & Eldholm, O. *Bull. geol. Soc. Am.* **88**, 969-999 (1977).
- Griffin, W. L. *et al. Q. Jl geol. Soc. Lond.* (in the press).
- Eldholm, O., Sundvor, E. & Myhre, A. *Mar. geophys. Res.* (in the press).
- Rønnevik, H. & Navrestad, T. *Geotjournal* **1**, 33-46 (1977).
- Lundquist, I. & Lagerbäck, R. *Geol. Förel. Stockh. Förel.* **98**, 45-51 (1976).
- Mørner, N.-A. *Geol. Förel. Stockh. Förel.* **99**, 48-54 (1977).
- Gabrielsen, R. H. & Ramberg, I. B. *Norsk. geol. Tidsskr.* (in the press).
- Beaumont, C. *Geophys. J. R. astr. Soc.* **55**, 471-497 (1978).

Recognition and initiation site for four late promoters of phage T7 is a 22-base pair DNA sequence

Nikos Panayotatos & Robert D. Wells

College of Agricultural and Life Sciences, Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

T7 late promoters have a 22-base pair sequence in common. The 22 base pairs are necessary and sufficient for recognition and initiation by T7 RNA polymerase.

T7 RNA POLYMERASE, the product of viral gene 1, is a single subunit protein of molecular weight 107,000 (refs 1, 2). This enzyme exhibits almost absolute specificity for the transcription

of the late T7 genes, both *in vivo* and *in vitro*^{3,4}. The early promoters for the *Escherichia coli* RNA polymerase are not recognised by the T7 enzyme and even the genes of the closely related phage T3 are transcribed only poorly⁵.

Transcription of intact T7 DNA with purified T7 RNA polymerase produces seven discrete transcripts initiated from promoter sites located in the 45-95% region^{3,4}. Additional promoter sites at 14.2, 14.6, 18.9, 22.6, 27.9 and, possibly, at

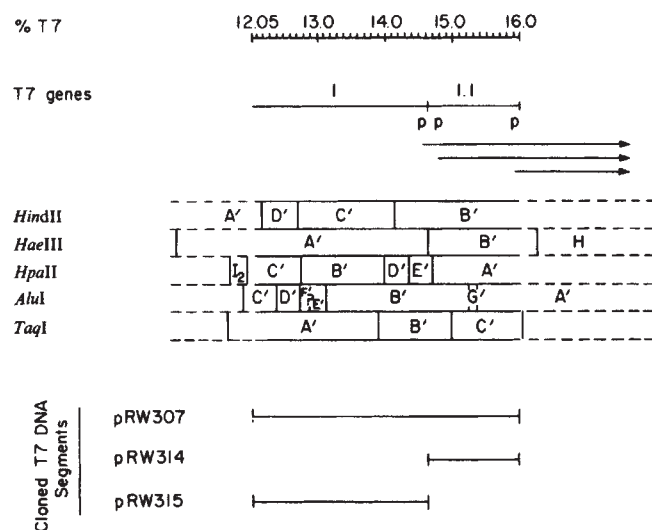


Fig. 1 Genetic and restriction map of a T7 DNA segment cloned in pRW307 (ref. 12). The restriction fragments inserted in the recombinant plasmids pRW314 and pRW315 are also indicated. The exact boundaries of the insertion in pRW307 have been determined with respect to the T7 map at 12.05% (less than 20 base pairs to the right of a *HpaI* site at 12.03%¹²) and at 16.00% (in the middle of the *HaeIII* site, as determined herein). The positions of the late promoters in this region at 14.6, 14.8 and 15.9% are indicated (p) and were identified in previous studies^{8,11}. The horizontal arrows point in the direction of transcription from each promoter. The restriction fragments of the vector which flank the insertion in pRW307 are designated by unprimed letters. The junctions between the solid and dashed lines represent the boundaries of the insertion. A more detailed map is presented elsewhere^{11,12}.

13.5% have been recognised from transcription studies on T7 DNA restriction fragments^{6,7}. The sequence of two restriction fragments containing the promoters at 14.6 and ~70% has been reported^{6,8}.

Recently, the presence of late promoters in cloned segments of the 12.05–18.19% region was investigated^{9–11}. The several hundred-fold stimulation of T7 RNA synthesis by some of the recombinant plasmids and the lack of stimulation by others narrowed the possible location of late promoters to three small regions^{10,11}. Subsequent determination of the number and size of the late transcripts synthesised off a variety of restriction fragments allowed the localisation of three late promoters inside these regions, at 14.6, 14.8 and 15.9% (ref. 11). These studies also demonstrated that late promoters were not located at 13.5 and 14.2%, in contrast with previous reports^{6,7}.

This article describes the sequence and template properties of two restriction fragments containing the promoters at 14.8 and 15.9%.

Isolation and sequencing of cloned promoter fragments

The T7 RNA polymerase promoters at 14.8 and 15.9% are located within *HaeIII* fragment B' which spans the 14.65–16.00% region cloned in pRW307 (refs 1, 2; Fig. 1). The promoter at 14.8% is also contained within a 112 base pair *HpaII*–*TaqI* fragment, whereas the promoter at 15.9% is contained within a 260 base pair *AluI*–*TaqI* fragment (Fig. 1). These smaller fragments were isolated and sequenced as described in Fig. 2. Autoradiograms of typical sequencing gels of the 112 and 260 base pair fragments are shown in Fig. 2a, b, and Fig. 2c, d, respectively. The complete sequence of the 112 base pair fragment is shown in Fig. 3a. The first 85 base pairs from the *TaqI* site are shown for the 260 base pair fragment in Fig. 3b.

On the basis of the restriction mapping data obtained for the 12.05–18.19% region of T7 DNA^{11,12}, the sequence of the 260 base pair fragment shown in Fig. 3b was expected to span the

region 15.8–16.0% of T7 DNA and extend past the insertion–vector junction in pRW307. (The exact end of the insertion was not known because the T7 DNA fragment had suffered a deletion during cloning¹².) The location of this junction was deduced from a comparison of the sequence in Fig. 3b with the published sequence of a 29-nucleotide RNA fragment transcribed from the 16.0% region of T7 DNA¹³. The sequence of the first 22 nucleotides from the 5' end of this RNA was identical with the sequence between nucleotides 50–71 of Fig. 3b. In contrast, the homology was lost over the next seven nucleotides. This abrupt loss of homology indicated that position 71 corresponded to the last base pair of the T7 DNA insertion before the vector junction in pRW307. The deletion of T7 DNA beyond nucleotide 71 caused the loss of the right half of a *HaeIII* site (GG↓CC) that is evident in the sequence of the RNA and has been mapped by F. W. Studier (personal communication) at 16.00% of the T7 genome. Thus, the restriction map of T7 DNA and the homology with the sequence of the short RNA are both consistent with the assignment of the end of the insertion at 16.00%.

The T7 fragment in pRW307 was inserted into the *HpaI* site of the pRT29 vector¹². The sequence of the vector-half of this site corresponds to nucleotides 72–74, indicating that the vector remained intact at this site, in agreement with the conclusions drawn from the restriction analysis data¹².

A recognition–initiation site for T7 RNA polymerase

Our previous studies have shown that two late promoters are located at 14.8 and 15.9% of the T7 genome¹¹, that is, within the sequences of the 112 and 260 base pair fragments shown in Fig. 3. Measurements of the lengths of the transcripts initiated at these promoters and terminated at various restriction sites downstream, placed the initiation points at ~93 and ~35 nucleotides to the left (upstream) of the *TaqI* sites of the 112 and 260 base pair fragments, respectively¹¹. The initiation points for these transcripts were more accurately determined from the data shown in Fig. 5 (lane a). The lengths of the transcripts shown in this figure were determined to be 93±4 and 47±2 nucleotides. These sizes place the initiation point at a corresponding number of base pairs upstream from the *TaqI* sites, within positions 16–24 of the 112 base pair fragment (Fig. 3a) and positions 37–41 of the 260 base pair fragment (Fig. 3b).

Inspection of the sequences around these two regions immediately revealed the presence of a nearly identical sequence of 23 contiguous nucleotides TAATACGACTCACTATAGGPuPuNA between positions 5–27 and 22–44 in Fig. 3a and b, respectively.

Comparison of this sequence with the published sequences of the promoters at 14.6 and ~70% (refs 6, 8) revealed almost complete homology, for the region around and immediately upstream from the transcription initiation site (Fig. 4). On the basis of these four structures, the following DNA sequence is proposed for the recognition–initiation site for T7 RNA polymerase:



In proposing this sequence, only nucleotides identical in all four promoters were considered, with the following two exceptions: (1) Position –5 of the promoter at 15.9% is a G rather than a C. We have assumed this transversion to be a feature of this particular promoter, and believe it to be responsible for the approximately fivefold lower activity observed relative to the promoters at 14.6 and 14.8% (ref. 11). (2) Position –16 of the promoter at ~70% reads C rather than A. Because the other three sequences have an A at this position, we have assumed that the A→C transversion is a deviation from the general structure. It may be of interest to determine the strength of this promoter relative to the others.

For consistency, we have chosen to include in the general structure the adenosine at position +6, although the continuity of the homology is interrupted at position +5. The preservation

Fig. 2 Autoradiograms of sequencing gels of two promoter-containing fragments of T7 DNA cloned in pRW307. *HaeIII* fragment B' (Fig. 1) was purified from a *HaeIII*+*EcoRI* digest of pRW307 by RPC-5 column chromatography¹⁴. This fragment was digested with *HpaII*, labelled at the 5' ends with T4 polynucleotide kinase and redigested with *TaqI*. These reactions produced a 23 base pair *HaeIII*-*HpaII* fragment, a 112 base pair *HpaII*-*TaqI* fragment containing the promoter at 14.8%, a 409 base pair *TaqI* fragment C' and a 99 base pair *TaqI*-*HaeIII* fragment. These fragments were fractionated on 8% polyacrylamide gels and visualized by autoradiography and ethidium staining. The 112 base pair fragment labelled at the 5' end of the *HpaII* site was recovered from the gel and sequenced. Lanes *a* and *b* show an autoradiogram of a typical sequencing gel for this fragment. The complete sequence of the *HpaII*-*TaqI* fragment which

spans the region 14.70–15.00% of T7 DNA is shown in Fig. 3a. The promoter at 15.9% is contained within a 260 base pair *AluI*–*TaqI* fragment which was isolated and sequenced as follows. The purified *HaeIII* fragment B' was digested with *TaqI*, labelled at the 5' ends with T4 polynucleotide kinase and redigested with *AluI*. These reactions produced a 135 base pair *HaeIII*–*TaqI* fragment, a 104 base pair *TaqI*–*AluI* fragment, a 45 base pair *AluI* fragment G', a 260 base pair *AluI*–*TaqI* fragment containing the promoter at 15.9%, and a 99 base pair *TaqI*–*HaeIII* fragment. These fragments were fractionated on 8% polyacrylamide gels and visualised by autoradiography and ethidium staining. The 260 base pair fragment (labelled at the *TaqI* site) was eluted from the gel and sequenced. An autoradiogram of a typical sequencing gel for this fragment is shown in lanes c and d. The sequence of the first 88 nucleotides from the *TaqI* site is shown in Fig. 3b. The sources of the enzymes and the procedures for the restriction reactions, end-labelling, gel elution and sequencing gel electrophoresis were based on published procedures¹⁵ and have been described^{11,12}.

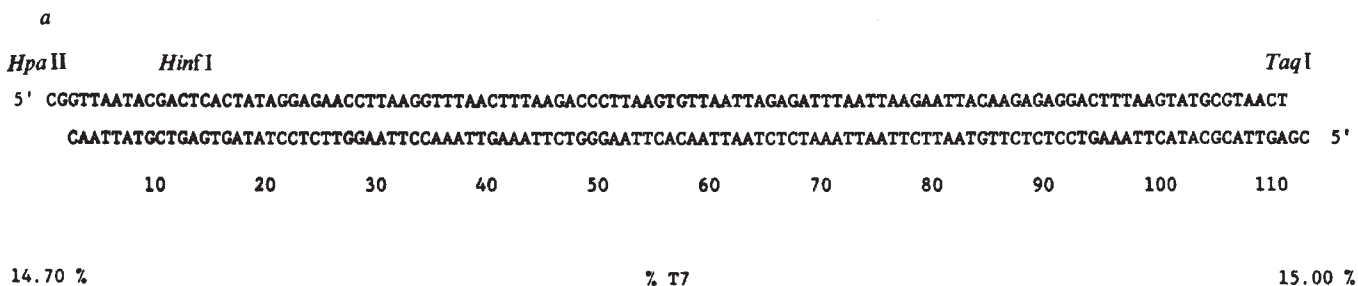
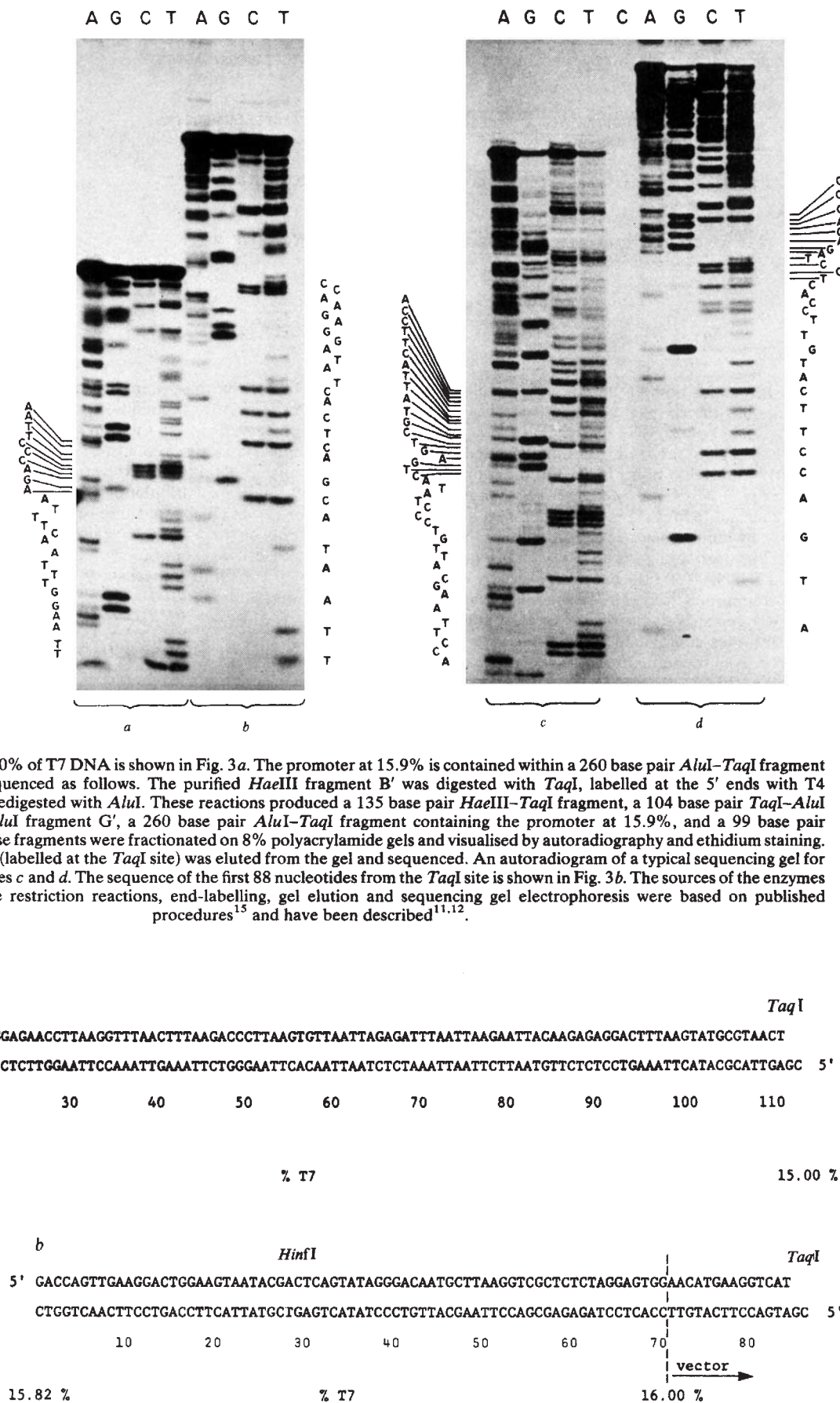
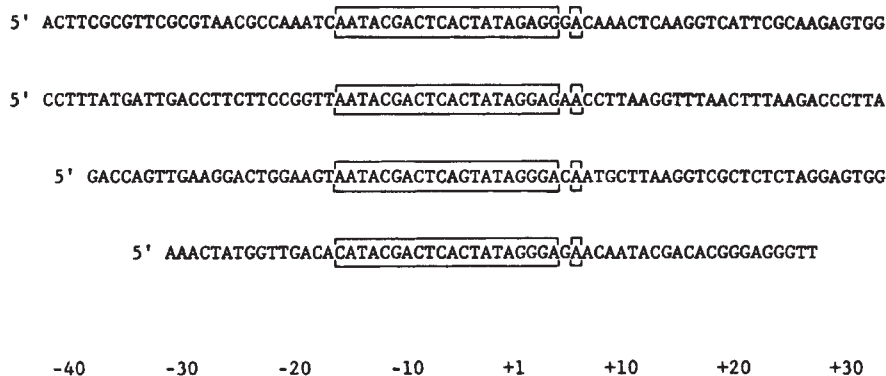


Fig. 3 DNA sequence of promoter-containing fragments. *a*, Nucleotide sequence of the 112 base pair *HpaII*-*TaqI* fragment spanning the region 14.70–15.00% of T7 DNA. This fragment was isolated from a larger cloned segment of T7 DNA and sequenced as

described in the legend to Fig. 2. *b*, Nucleotide sequence of part of the 260 base pair *AluI*-*TaqI* fragment showing the region 15.82–16.00% of T7 DNA and 17 base pair of vector DNA at the *TaqI* end. This fragment was isolated from a longer, cloned segment of T7 DNA and sequenced as described in Fig. 2. Positions 70–71 correspond to the vector–insertion junction at 16.00% of T7 DNA. Positions 69–73 cover half of the *HaeIII* site of T7 and half of the *HpaI* site of the vector (see text).





of adenosine at position +6 may not be fortuitous but may be due to an important contact of the polymerase with this nucleotide.

The DNA region between positions -16 and +6 around the transcription initiation point is proposed to constitute a recognition site for T7 RNA polymerase because of its highly preserved sequence in the four promoters examined thus far. Random occurrence of the common sequence in T7 DNA is extremely unlikely. (The probability for a 16 base pair sequence to occur 4 times in T7 DNA (40,000 base pairs) is 8×10^{-16} (W. Fitch, personal communication). The probability for the 22 base pair common sequence is even lower but calculation of the exact number becomes complicated by the possible variation in the PuPuPuN part of the sequence.) However, this region might not contain all pertinent structural elements. If nucleotides outside

the common sequence also contributed structural information necessary for recognition, their positions should also have been preserved among the four sequences. Inspection of the nucleotides outside the common sequence (Fig. 4) shows that the same nucleotide does not appear at a given position more than twice, with the exception of the adenosine at -19 and the thymidine at +21, which appear three times. This complete lack of homology outside the -16 to +6 region, as opposed to the almost complete preservation of homology inside, suggests that positions -16 to +6 constitute a necessary and sufficient recognition-initiation site.

Further identification of the promoter

Direct evidence to support the previous conclusion comes from the results of cloning experiments. In pRW315, the T7 DNA region beyond the *Hae*III site at position +33 of the 14.6% promoter was substituted by vector DNA. In pRW314, the T7 DNA region ahead of the *Hae*III site at position -42 of the 14.8% promoter was also substituted by vector DNA (Figs 1 and 4). Despite the substitution of the natural viral sequences by vector DNA, the activity of these promoters in pRW314 and pRW315 was not significantly reduced relative to their activity in pRW307 (ref. 11).

Similar conclusions were reached when purified *Hae*III fragments A' and B' were assayed¹¹. In both cases, discrete transcripts were obtained at comparable yields from the promoters at 14.6 and 14.8%, indicating that the absence of DNA upstream from position -42 and downstream from position +33 from the initiation point was not required by T7 RNA polymerase.

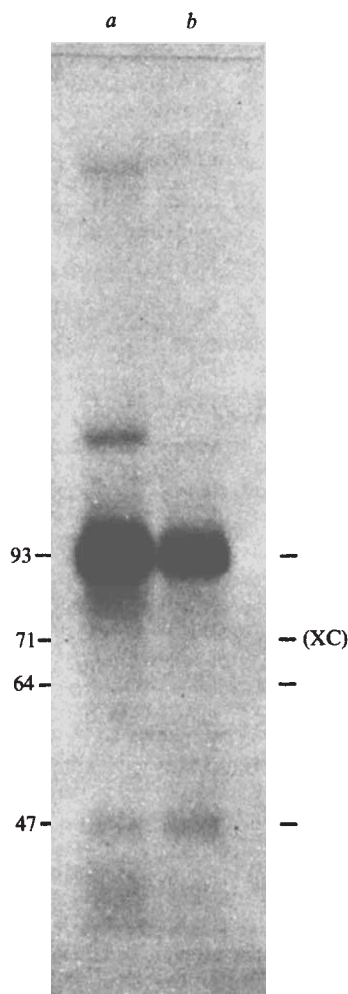


Fig. 5 Autoradiogram of ^{32}P -labelled transcripts fractionated on a 8% polyacrylamide-7 M urea gel. RNA was synthesised with highly purified T7 RNA polymerase (0.15 μg of fraction V; the gift of G. Kassavetis and M. Chamberlin⁷) as described previously¹¹; 0.75 μg of either *Hae*III fragment B' digested with *Taq*I (lane a) or *Hae*III fragment B' digested with *Taq*I + *Hpa*II (lane b) was used as the template. The heaviest bands in both lanes a and b correspond to 93 ± 4 base transcripts¹¹ initiated at the 14.8% promoter and terminated at the *Taq*I site. The weak bands near the bottom of the gel correspond to 47 ± 2 base long transcripts initiated at the 15.9% promoter and terminated at a *Taq*I site. The $\sim 5 \times$ weaker expression of the 15.9% promoter relative to the 14.8% promoter (as judged by band intensity) is consistent with previous observations¹¹ and is attributed to a variation in the structure of the 15.9% promoter (see text). The weak band in the upper half of the gel in lane a apparently represents an end-to-end transcript synthesised off the 135 base pair *Hae*III-*Taq*I fragment. When this fragment is cleaved by *Hpa*II 23 base pairs from the *Hae*III site upstream from the 14.8% promoter, this band disappears (lane b). A 64 base transcript synthesised by *E. coli* RNA polymerase off a 203 base pair *Hae*III fragment containing the *lac* UV5 promoter was also used as a marker²⁰. Its position, as well as the position of xylene cyanol which migrates as a 71-base long transcript on this gel, is indicated.

Fig. 4 DNA sequence around the T7 late promoters located at 14.6 (ref. 6), 14.8 (this report), 15.9 (this report) and $\sim 70\%$ (ref. 8) of the genome. The highly preserved sequence which constitutes the recognition-initiation site is indicated by the solid lines. Positions are marked from the initiation point at +1. The complete DNA sequence between the promoters at 14.6 and 14.8% is shown on the top two lines, interrupted at the *Hae*III site (at 14.65%) for clarity.

Further evidence on the boundaries of the sequence that defines a T7 late promoter was obtained from transcription studies (Fig. 5) of the 112 base pair *HpaII*-*TaqI* fragment (at 14.70–15.00% in Fig. 1) in which the DNA sequence ends just 4 base pairs upstream from the recognition site (Fig. 3a). The activity of the 14.8% promoter in this fragment was determined relative to the 135 base pair *HaeIII*-*TaqI* fragment (at 14.65–15.00%) in which the T7 DNA extends through the *HpaII* site upstream to the *HaeIII* site (Fig. 1). Each reaction mixture also contained *TaqI* fragment C' (Fig. 1); thus, the expression of the 15.9% promoter in the latter fragment served as an internal control. In identical assay conditions, comparable amounts of discrete transcripts (~93 bases long) were synthesised from the 14.8% promoter in either the *HpaII*-*TaqI* or the *HaeIII*-*TaqI* fragments. In both cases, the 47-base transcript from the 15.9% promoter control was expressed to a similar extent. This result demonstrated that elimination of the DNA as close as 4 base pairs upstream from the common promoter sequence had no effect on the efficiency or selectivity of transcription.

In summary, the lack of homology outside the 22 base pair common sequence indicates that the neighbouring DNA structure did not contribute directly to recognition. The results of experiments in which the DNA upstream or downstream from the common sequence was either replaced by foreign DNA or removed are consistent with this interpretation. By these criteria, the 22 base pair contiguous sequence constitutes a fully defined recognition-initiation site for T7 RNA polymerase.

Comparison with other promoters

The sequence of the recognition site contains very little two-fold symmetry. Only the 6 base pair sequence between positions -5 and +1 is symmetrical and additional symmetry around the same axis depends on the particular nature of the nucleotides in the -PuPuPuN- part of the sequence. In this regard, the 14.6% promoter contains much more symmetry than the 14.8% promoter, but this feature does not seem to affect the relative expression of these two promoters. The lack of extensive two-fold symmetry is also observed with bacterial promoters¹⁶.

In promoters for *E. coli* RNA polymerase, a 7 base pair site of partially preserved sequence (the 'Pribnow box') is located 7 base pairs upstream from the transcription initiation point¹⁷. A second region of partially preserved sequence also exists approximately 30 base pairs further upstream and mutations in this region affect promoter function^{16,18}. In contrast, T7 promoters consist of one contiguous recognition site, which

includes the initiation point and no other sequence further upstream, seems to be required for complete function. In this respect, T7 promoters are quite distinct from their bacterial counterparts. The physical separation of the two recognition sites and the initiation site in the *E. coli* promoters, as opposed to the condensation of these elements into one contiguous site in T7 promoters, may reflect the size difference between the two enzymes (the *E. coli* polymerase is about five times larger than the T7 RNA polymerase).

The possibility has been raised that bacterial promoters may be located at easily denatured, A+T-rich regions of the DNA template¹⁹. Such a correlation was also proposed for T7 late promoters⁵. The unusually high (69%) A+T content of the 112 base pair fragment containing the promoter at 14.8% agrees with this model. However, neither the promoter at 14.6% nor the promoter at 15.9% are located in unusually A+T rich regions. (A+T = 52% for both *HpaII* fragment E and the 71 base pair T7 segment around 15.9% in Fig. 3b.) Thus, A+T richness does not seem to be a general property of T7 promoter sites. The unusually high A+T content of the 14.7–15.00% region may be important for some other regulatory site in the vicinity of this region, such as the origin of T7 DNA replication¹².

Noted added in proof: M. Rosa²¹ has recently reported that four class 3 T7 late promoters share the same sequence.

Received 30 March; accepted 21 May 1979.

1. Chamberlin, M., McGrath, J. & Waskell, L. *Nature* **288**, 227–231 (1970).
2. Summers, W. C. & Siegel, R. B. *Nature* **228**, 1160–1162 (1970).
3. Golomb, M. & Chamberlin, M. *J. biol. Chem.* **249**, 2858–2863 (1974).
4. Niles, E. G. & Condit, R. C. *J. molec. Biol.* **98**, 57–67 (1975).
5. Golomb, M. & Chamberlin, M. *Proc. natn. Acad. Sci. U.S.A.* **71**, 760–764 (1974).
6. Oakley, J. L. & Coleman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4266–4270 (1977).
7. Kassavetis, G. A. & Chamberlin, M. J. *J. Virol.* **29**, 196–208 (1979).
8. Oakley, J. L., Strothkamp, R. E., Sarris, A. H. & Coleman, J. E. *Biochemistry* **18**, 528–537 (1979).
9. Panayotatos, N. *Fedn Proc.* **37**, 1409 (1978).
10. Panayotatos, N. & Wells, R. D. *Abstract to XI International Congress of Biochemistry*, Toronto, Canada, July 8–13 (1979).
11. Panayotatos, N. & Wells, R. D. (in preparation).
12. Panayotatos, N. & Wells, R. D. *J. biol. Chem.* (in the press).
13. Robertson, H. D., Dickson, E. & Dunn, J. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 822–826 (1977).
14. Wells, R. D. *et al. Meth. Enzym.* (in the press).
15. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
16. Scherer, G. E. F., Walkinshaw, M. D. & Arnott, S. *Nucleic Acids Res.* **5**, 3759–3773 (1978).
17. Pribnow, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 784–788 (1975).
18. Gilbert, W. in *RNA Polymerase* (ed. Losick, R. & Chamberlin, M.) 193–205 (Cold Spring Harbor Laboratory, 1976).
19. Chamberlin, M. J. *A. Rev. Biochem.* **43**, 721–775 (1974).
20. Maquat, L. E. & Reznikoff, W. S. *J. molec. Biol.* **125**, 467–490 (1979).
21. Rosa, M. *Cell* **16**, 815–825 (1979).

letters

On the detection of large-scale inhomogeneities in the Universe

THE detection of large-scale inhomogeneities in the Universe has proceeded along a number of fronts. The observed spatial distribution of galaxies indicates clumping of matter on length scales up to ~30 Mpc (ref. 1). Limits on fine-scale temperature fluctuations in the microwave background² constrain mass concentrations at the epoch of recombination, and later^{3,4}. Neither of these methods is particularly sensitive to large-scale low-density contrast ($\delta\rho/\rho \ll 1$), perturbations at the current epoch (that is, masses $\gg 10^{19} M_{\odot}$ on a scale $\geq 1,000$ Mpc), although in principle the distribution of discrete radio sources on the sky may be investigated for such effects (see for example ref. 5). We show here that measurements of the isotropy of the X-ray

background provide a sensitive means of detecting such large-scale inhomogeneities within $z \sim 5$.

A matter perturbation, with excess mass δM , retards the motion of our Galaxy relative to the overall expansion of the Universe with a peculiar velocity δv_G given, from energy considerations, by

$$v_G \delta v_G \approx \frac{G \delta M}{r} \quad (1)$$

Here the perturbation is at a distance r from our Galaxy and v_G is the Hubble flow velocity over that distance. We then obtain

$$\delta v_G \approx \frac{1}{2} \left(\frac{\delta\rho}{\rho} \right) \Omega H_0 \frac{l^3}{r^2} \quad (2)$$

where Ω is the cosmological density parameter ($8\pi G\rho/3H_0^2$),

$\delta\rho$ is the excess matter density and l is the size of the perturbation.

Other cosmological effects are small, and can be neglected, provided that $(\delta\rho/\rho) \ll 1$ and r corresponds to a redshift $\lesssim 1$. The peculiar velocity of the Galaxy then induces a 24-h variation⁶ in the microwave background through the Compton-Getting effect;

$$I = (2 - \alpha) \left(1 + \frac{\delta v_G}{c} \cos \theta \right) I_0 \quad (3)$$

where θ is measured from the direction towards δM , and α is the photon number spectral index of the radiation. ($\alpha = 1$ for the Rayleigh-Jeans part of the microwave spectrum and ~ 1.4 for the X-ray spectrum; $2 < E < 20$ keV.) Twenty-four-hour anisotropy in the microwave background, implying a peculiar velocity of 600 km s^{-1} , has been reported by Smoot *et al.*⁷ after correcting for the rotation of the Galaxy.

The microwave background originates at $z \sim 10^3$, whereas the X-ray background probably originates from within $z \lesssim 3$. Only weak evolution of the known classes of X-ray source⁸ (particularly quasars and active galaxies) is needed to explain the observed intensity. If $r \ll cH_0^{-1}$ (that is if the redshift of the matter perturbation $z_M \ll 1$), then a similar 24-h variation may be expected in both the X-ray and microwave backgrounds. On the other hand, if $r \sim cH_0^{-1}$, δM generates large-scale shear across the Universe out to $z \sim 1$ and the X-ray background varies in a significantly different manner. A 12-h variation, peaking at $\theta \sim 90^\circ$, is predicted as shown in Fig. 1. The sources of the X-ray background between ourselves and the mass are decelerating more rapidly than we are and are thus seen to be receding. In the opposite direction we are decelerating more than the background, which also gives a recession. Only in intermediate directions is our velocity and that of the background comparable. The 12-h effect reduces if r is much greater than cH_0^{-1} . The peculiar velocity δv_r of the Galaxy relative to a shell of X-ray emitting matter at distance d is given, from (2) as

$$\delta v_r = \delta v_G \left(\cos \theta + \frac{(d/r - \cos \theta)}{(1 + (d/r)^2 - 2(d/r) \cos \theta)^{3/2}} \right) \quad (4)$$

This gives a 12-h variation through equation (3) with $\delta v_G \cos \theta$ replaced by δv_r . The overall effect on the X-ray background may be further complicated by any X-ray emission from δM itself, and by the rotation of the Galaxy together with any other local peculiar velocities

An oblate 12-h variation in the X-ray background can thus, in principle, measure large-scale shear over distances out to $z \sim 1$. The net differential velocities between different directions need not exceed more than a few hundred km s^{-1} before giving a

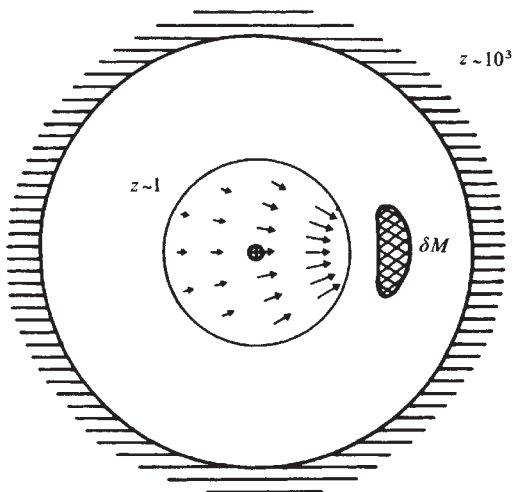


Fig. 1 Schematic representation of tidal gravitational influence of density inhomogeneity, mass δM , on the velocity shear of the X-ray background.

detectable X-ray signal. This is a consequence of the fact that the X-ray background provides an absolute measurement of integrated quantities out to $z \sim 1$. This has been noted by Wolfe⁹ in constraining world models with shear using the 1% limit to 12- and 24-h variations in the X-ray background placed by Schwartz¹⁰. The Leicester Sky Survey Instrument on Ariel V can detect such variations at the level of $\sim 0.5\%$ (ref. 11).

The 12-h isotropy of the X-ray background is most sensitive to masses situated at $z_M \sim 1$;

$$\delta M = 10^{20} z_M^2 H_{50}^{-1} \delta v_6 \quad M_\odot$$

with size

$$l = 1,000 \left(\frac{\delta\rho}{\rho} \right)^{-1/3} z_M^{2/3} H_{50}^{-1} \delta v_6^{1/3} \Omega^{-1/3} \quad \text{Mpc}$$

where $\delta v_G = 600 \delta v_6 \text{ km s}^{-1}$, $H_0 = 50 H_{50} \text{ km s}^{-1} \text{ Mpc}^{-1}$.

A large-scale density inhomogeneity, with $(\delta\rho/\rho) \sim 0.01$ now, corresponds to $(\delta\rho/\rho) \sim 10^{-5}$ at the epoch of recombination. We conclude that the combination of microwave and X-ray background anisotropy observations will enable us to study the very large scale structure of the Universe at the current epoch.

We thank Professor Martin Rees for discussions. A.C.F. thanks the Radcliffe Trust for support.

A. C. FABIAN

*Institute of Astronomy,
Madingley Road,
Cambridge, UK*

R. S. WARWICK

*Physics Department,
University of Leicester,
Leicester, UK*

Received 4 April; accepted 30 May 1979.

1. Peebles, P. J. E. *IAU Symp.* 79, 217-225 (1978).
2. Boynton, P. E. *IAU Symp.* 79, 317-324 (1978).
3. Rees, M. J. & Sciama, D. W. *Nature* 217, 511-516 (1968).
4. Dyer, C. C., *Mon. Not. R. astr. Soc.* 175, 429-447 (1976).
5. Webster, A., *Mon. Not. R. astr. Soc.* 175, 71-84 (1976).
6. Peebles, P. J. E. & Wilkinson, D. T. *Phys. Rev.* 174, 2168 (1968).
7. Smoot, G. F., Gorenstein, M. V. & Müller, R. A., *Phys. Rev. Lett.* 39, 898-901 (1977).
8. Rowan-Robinson, M. & Fabian, A. C. *Mon. Not. R. astr. Soc.* 170, 199 (1975).
9. Wolfe, A. M. *Astrophys. J. Lett.* 159, L61-L67 (1970).
10. Schwartz, D. *Astrophys. J.* 162, 439-444 (1970).
11. Warwick, R. S., Pye, J. P. & Fabian, A. C. (in preparation).

The possibility of accretion onto radio pulsars

AS an explanation for the observed long-term variations in pulsar emission¹⁻³, an accretion model is proposed here which permits interstellar neutral hydrogen to infiltrate the light-cylinder and, following ionisation, to modulate the pulsed emission. This would suggest that pulsars have an eventful passage through the interstellar medium, and do not always remain as pulsars.

The widely accepted picture of a pulsar is that of a magnetised neutron star rotating in a vacuum⁴. The intense electric fields thereby induced are supposed to remove charged particles from the surface near the magnetic poles and accelerate them to high energies⁵. The radiation generated by these particles then interacts with the local magnetic field to produce cascades of electron-positron pairs^{6,7}, whose emission we observe. The site of the emission is still unclear. Around the light-cylinder the rotating magnetic field produces a strong electromagnetic wave whose pressure prevents the interstellar medium from moving into the magnetosphere. In all emission models based on this general picture it is vital that the region inside the light-cylinder be maintained as a near-vacuum.

Theoretical attention is usually concentrated on explaining emission features which have time scales of the order of the neutron star rotation period or less^{6,8} (microstructure⁹, for example). Instabilities of the flow between the magnetic poles

and the light-cylinder are seen as a natural explanation for such short time-scale modulations. It has been observed, however, that both individual and integrated pulse intensities are also strongly modulated on every time scale from hours to years^{1,2}. Recently, substantial day-to-day changes in the spectra of nine pulsars have been reported³. It is difficult to see how such long-term variability can be achieved by any mechanism whose fundamental time scale is the period of rotation. On the other hand, the shape of the integrated profiles and their polarisation are generally stable¹⁰ (although even here occasional 'mode-changing' occurs¹¹). The overall impression is that a basic emission pattern is being modulated by some external influence.

The Solar System was once thought to be insulated from the interstellar medium by the shielding effect of the solar wind¹². Later work has shown that the Solar System is in fact penetrated by interstellar hydrogen atoms¹³. This is made possible by the relative motion of the Sun and the interstellar medium (about 15 km s⁻¹). One is led to speculate whether pulsars, whose velocities are often well in excess of 100 km s⁻¹, may be subject to such a process.

The radius, R_w , of a pulsar's wave-zone is given by the condition

$$\frac{\dot{E}}{4\pi R_w^2 c} \sim \frac{B_0^2}{8\pi}$$

where $\dot{E}$ is the spin-down energy loss-rate and B_0 is the ambient interstellar magnetic field. Taking $\dot{E} \sim 10^{31}$ erg s⁻¹ and $B_0 \sim 5 \times 10^{-6}$ G, we obtain $R_w \sim 5 \times 10^{15}$ cm. By comparison, the mean free path of a hydrogen atom with respect to ion collisions is $\lambda \sim 1/(n_i \sigma)$, where n_i is the ion density and σ the collisional cross-section. Taking $n_i \sim 0.03$ cm⁻³ and $\sigma \sim 3 \times 10^{-15}$ cm² we obtain^{14,17} $\lambda \sim 10^{16}$ cm. Thus $\lambda > R_w$ at least for weaker (low period change $\dot{P}$) pulsars and, although ions will be deflected by the strong wave, neutral hydrogen atoms can pass insensibly through the interface between the interstellar medium and the wave and move on hyperbolic orbits towards the neutron star.

The highly energetic particles emitted by the pulsar present no threat to the oncoming atoms, both because of their extremely low cross-sections ($\sim 10^{-18}$ cm²) and low density ($\ll 1$ cm⁻³) outside the light-cylinder. But thermal UV radiation from the star's surface may be more dangerous. It can be shown that an atom travelling at 100 km s⁻¹ will survive against photoionisation until it enters the light-cylinder radius ($\sim 5 \times 10^9$ cm) only if the surface temperature has fallen below 10⁵ K. On favourable assumptions, this temperature may be reached within 0.1 Myr of the formation of the neutron star¹⁵. Similar conclusions have been reached independently by Tsygan¹⁶.

If an atom is ionised before reaching the light-cylinder, it will be driven out of the magnetosphere by the pulsar's strong wave. But if ionisation takes place within the light-cylinder (through interaction either with the thermal or the beamed radiation) it can play a role similar to that ascribed in some models^{6,7} to pair-creation: one of the two resulting charged particles will, if on an open field-line, be pulled to the polar cap by the electric field, whilst the other will be accelerated outwards and expelled. If charged particles can be supplied in sufficient numbers, the electric field parallel to the magnetic field-lines will be 'short-circuited'. This would result in the breakdown of any emission mechanism which requires particles to be removed from the surface.

For a pulsar of period P moving with a velocity V through a medium of density n_H the rate of accretion of neutral particles into the light-cylinder radius, R_L , is $\pi R_L R_A V n_H$, where R_A is the 'accretion radius' of the neutron star defined by $R_A = 2GM/V^2$. This is the appropriate formula when $\lambda \gg R_A$. Taking $M \sim 1.4 M_\odot$, $V \sim 100$ V₁₀₀ km s⁻¹, we obtain the accretion rate

$$\dot{N}_A \sim 5.6 \times 10^{29} \frac{P n_H}{V_{100}} \text{ s}^{-1}$$

To eliminate the electric field along the magnetic fieldlines we require a charged particle flux of $(\pi R_p^2)(B/\text{Pec})c$, where B is the polar magnetic field strength and R_p the cap radius (whose circumference is defined by the fieldlines which just touch the

light-cylinder). With $R_p \sim 1.4 \times 10^4/P^{1/2}$ cm and $B \sim 10^{12}$ G the necessary particle flux is

$$\dot{N}_p \sim \frac{1.3 \times 10^{30}}{P^2} \text{ s}^{-1}$$

For the typical values $P \sim 1$ s and $n_H \sim 1$ cm⁻³, $\dot{N}_A$ is a significant fraction of $\dot{N}_p$ and disturbance of the emission pattern can be expected. Comparing the expressions for $\dot{N}_A$ and $\dot{N}_p$, we see that this effect is more likely to be pronounced in high-period, low-velocity pulsars. A further consequence is that fast-moving pulsars (through a lower accretion rate) may actually 'live' longer than slow-moving pulsars.

Long-term variations in emission can be interpreted, through variations in n_H , as the result of inhomogeneities in the interstellar medium. It is therefore possible to regard pulsars as monitors of the detailed structure of the gas in their neighbourhood. Furthermore, one might question the usual anthropomorphic assumption of a single uninterrupted lifespan for each pulsar. It is believed¹⁷ that the Solar System has passed through many neutral hydrogen clouds in the 5×10^3 Myr since its formation. In its much shorter life (~ 2 Myr) a fast-moving pulsar will also have led a chequered existence, passing through clouds of various densities. The calculations here suggest that even a 'standard' H I cloud¹⁸ ($n_H \geq 10$ cm⁻³) can switch a pulsar off. A pulsar moving at 100 km s⁻¹ covers 200 pc in 2 Myr. On current estimates¹⁷, 12 pc of this (or 6%) will be spent in clouds with $n_H \geq 10$ cm⁻³. There are now over 300 known pulsars, so there might be 20 single neutron stars which would otherwise have been detectable as pulsars at present immersed in dense H I regions. Pulsars are probably born in or near OB associations, so they are more likely to be surrounded by gas in their early lives and it may be significant that no pulsar has yet been found in a supernova remnant shell.

Accretion onto neutron stars suggests affinities between the model given here and the so-called X-ray 'bursters' (especially those which have yet to reveal evidence of being in a binary system¹⁹). These are often of a transient nature, inviting comparison with the 'nulling' phenomenon of radio pulsars, whereby emission ceases abruptly at all wavelengths. The time spent in the null state varies from pulsar to pulsar, but values of over 50% have been given for some pulsars²⁰. There is evidence²⁰ that nulling is more common in long-period pulsars, as would be expected from the above accretion rate estimates ('drifting subpulses'²¹, which are often associated with nulling, may therefore also be related to accretion). The time spent in a null sequence might sometimes be related to the energy of the following or preceding stretches of emission (as in MXB1730-335)¹⁹.

I thank H. J. Fahr, M. Walmsley, P. Biermann and the members of the Bonn pulsar group for helpful discussions.

G. A. E. WRIGHT

*Astronomische Institute der Universität Bonn,
Auf dem Hügel 71,
D-5300 Bonn, FRG*

Received 19 April; accepted 23 May 1979.

- McLean, A. I. O. *Mon. Not. R. astr. Soc.* **165**, 133 (1973).
- Huguenin, G. R., Taylor, J. H. & Helland, D. J. *Astrophys. J.* **181**, L139 (1973).
- Kuzmin, A. D. et al. *Mon. Not. R. astr. Soc.* **185**, 441 (1978).
- Gold, T. *Nature* **218**, 731 (1968).
- Goldreich, P. & Julian, W. H. *Astrophys. J.* **157**, 869 (1969).
- Ruderman, M. A. & Sutherland, P. G. *Astrophys. J.* **196**, 51 (1975).
- Sturrock, P. A. *Astrophys. J.* **164**, 529 (1971).
- Benford, G. & Buschauer, R. *Mon. Not. R. astr. Soc.* **179**, 189 (1977).
- Cordes, J. M. & Hankins, T. H. *Nature* (in the press).
- Manchester, R. N. & Taylor, J. H. *Pulsars*, 30 (Freeman, San Francisco 1977).
- Backer, D. C. *Nature* **228**, 1297 (1970).
- Axford, W. I., Dessler, A. J. & Gottlieb, B. *Astrophys. J.* **137**, 1268 (1963).
- Fahr, H. J. *Astrophys. Space Sci.* **2**, 474 (1968).
- Lang, K. R. *Astrophysical Formulae* (Springer, Berlin, 1974).
- Tsuruta, S., Canuto, V., Lodenquai, J. & Ruderman, M. A. *Astrophys. J.* **176**, 739 (1972).
- Tsygan, A. I. *Sov. astr. Lett.* **3**, 289 (1977).
- Talbot, R. J. & Newman, M. J. *Astrophys. J. Suppl.* **34**, 295 (1977).
- Spitzer, L. Jr *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).
- Lewin, W. H. G. *Adv. Space Explor.* (in the press).
- Ritchings, R. T. *Mon. Not. R. astr. Soc.* **176**, 249 (1976).
- Unwin, S. C., Readhead, A. C. S., Wilkinson, P. N. & Ewing, M. S. *Mon. Not. R. astr. Soc.* **182**, 711 (1978).

The size of Jupiter's electrically conducting fluid core

WHEN the magnetic field of a planet is due to hydromagnetic dynamo action in an electrically conducting fluid core surrounded by a poorly conducting mantle, Hide's¹ method can, in principle, be used to determine the radius r_c of the core from determinations of secular changes in the magnetic field $\mathbf{B}$ in the accessible region above the surface of the planet, mean radius r_s ($\geq r_c$). It will be possible to apply this method to Jupiter (and other planets) when magnetic measurements of sufficient accuracy and detail become available. We describe here a preliminary study in which we have analysed the magnetic measurements made in December 1973 and December 1974 when the Pioneer 10 and Pioneer 11 fly-by space-probes encountered Jupiter². We expected that over such a short time interval any true secular changes would be masked by errors and the corresponding estimate of r_c/r_s highly implausible or even physically impossible, but this turns out not to be the case. Taken at their face value the apparent secular changes in the dipole and quadrupole components of Jupiter's magnetic field imply that r_c/r_s is close to 0.7. Somewhat higher values of r_c/r_s are found when contributions from the octupole component are also included.

Jupiter is the largest known planet. It has a mean radius r_s of 6.97×10^7 m, more than 10 times that of the Earth, and a mean density of $1,330 \text{ kg m}^{-3}$, close to that of the Sun and about a quarter of that of the Earth. At about 5×10^{12} Pa (50 Mbar), Jupiter's central pressure is more than 10 times that of the Earth and this ensures that Jupiter's main chemical constituent, hydrogen, is metallic throughout most of the planet. Only above the level $r = r_m$ (r being the distance from the centre of the planet), in the 'mantle' and the 'atmosphere' where the prevailing pressures are less than a certain pressure of a few megabars, does hydrogen take its more familiar molecular form. This transition pressure is not yet known to much better than a factor of two and the corresponding estimates of r_m/r_s given in the recent literature range from 0.7 to 0.8. The planet is generally considered to be fluid throughout²⁻⁴.

The existence of a strong jovian magnetic field of internal origin and nearly dipolar form, with a dipole moment about 10^4 times the magnitude of and opposite in sign to that of the Earth's present field, was first inferred in the 1950s from radio-astronomical observations at decametre and decimetre wavelengths⁵. Important details of the field were obtained by Smith, Davis and Jones, and by Acuña and Ness from *in situ* magnetometer measurements on board the Pioneer 10 and Pioneer 11 space probes² and further details are expected from measurements on board Voyager 1 spacecraft, which passed close to Jupiter very recently (1–5 March 1979) and Voyager 2, which will encounter Jupiter later this year (2–9 July).

The only serious suggestion as to the origin of Jupiter's magnetic field closely parallels ideas developed by geophysicists over the past 30 years towards an explanation of the Earth's magnetism^{2,5}. Thus, we can suppose that the jovian field is due to ordinary electrical currents circulating in the conducting regions within the planet and that these currents are produced by the hydromagnetic dynamo process, involving inductive interactions between fluid motions and the magnetic field. Whether dynamo action is possible depends *inter alia* on the value of the so-called magnetic Reynolds number

$$R = \sigma \mu L U \quad (1)$$

where R is the ratio between the time scale of free ohmic decay $\sigma \mu L^2$ and an advective time scale L/U ; L is a typical length scale; U is a typical flow speed associated with the fluid motions, and σ and μ are average values of the electrical conductivity and magnetic permeability of the electrically conducting core of outer radius r_c within which dynamo action is hypothesised to occur. Such action is impossible when R is very small for then

motional induction cannot overcome the effects of ohmic dissipation, and also when R is infinite as the magnetic flux linkage of a perfect electrical conductor cannot change. The most efficient dynamos occur when R has some optimum value $\bar{R}$ where typical values of $\bar{R}$ range between 10 and 100.

It is tempting to assume that r_c is equal to r_m , the radius of the metallic core, as it would be if the mantle of Jupiter were a very good insulator. But the value of σ required to make $R = \bar{R}$ might be a good deal less than the high values (typically greater than $10^6 \Omega^{-1} \text{ m}^{-1}$) found in the metallic core, and it is conceivable that impurities and high pressure and temperature render the lower reaches (at least) of the molecular hydrogen mantle sufficiently conducting to support dynamo action there, in which case r_c would be significantly greater than r_m (refs 2, 4, 5).

The new method^{1,6} which estimates r_c with a typical error of the order of but greater than $r_c/\bar{R}$, has been applied to the Earth, using the best determination of the geomagnetic secular variation for the period 1955–75 expressed in terms of a spherical harmonic series truncated at degree $n = n^* = 8$. Values of r_c were obtained differing by typically 10% or 15% from the radius of the liquid metallic core as determined from seismological data. (The agreement to within 2% for the first epoch considered has been checked and confirmed but the subsequent analysis of data from other epochs suggests that 2% is atypically low⁷.) When sufficiently accurate measurements of the magnetic fields of other planets (for which seismological data are not available) have been made, it should be possible to use the method in the investigation of their interiors. The Pioneer 10 and Pioneer 11 measurements of Jupiter's magnetic field² are not, however, considered to be sufficiently accurate and detailed to give reliable information on secular changes in $\mathbf{B}$ during the short interval of only one year between the times $t = t_{10}$ and $t = t_{11}$ (say) of the two encounters². We expected that these uncertainties would manifest themselves as highly improbable or even physically impossible values of r_c based on these limited measurements but surprisingly we found this was not the case.

Smith *et al.*² have constructed several spherical harmonic models of the jovian magnetic field for $t = t_{10}$ and $t = t_{11}$. These are based on the vector helium magnetometer observations on board Pioneer 10 and Pioneer 11 made within $8r_s$ of the centre of Jupiter, in the region of the inner magnetosphere where the contribution to $\mathbf{B}$ made by local currents should be negligible. For the inner magnetosphere $\mathbf{B}$ consists of an 'internal' part originating within the planet and a smaller 'external' part due to electrical currents flowing in the middle and outer magnetosphere or in the magnetopause. Smith *et al.*² give for each of the Pioneer 10 and Pioneer 11 data sets four spherical harmonic models, each comprising sets of coefficients for the field of internal origin up to degree and order n_i and for the field of external origin coefficients up to degree and order n_e , as follows: (a) $n_i = 3, n_e = 2$; (b) $n_i = 3, n_e = 1$; (c) $n_i = 3, n_e = 0$; (d) $n_i = 2, n_e = 2$. For each of the four pairs of models we have calculated the total 'unsigned' flux through spheres of various radii $r = r_s(1 - 0.05p)$, $p = 0, 1, 2, 3, 4$, and so on, by numerically approximating the integral

$$N(t; r) = \int_0^{2\pi} \int_0^\pi |B_r(r, \theta, \lambda, t)| r^2 \sin \theta \, d\theta \, d\lambda \quad (2)$$

from values of B_r synthesised on a 10° by 10° zenographic θ (co-latitude) and λ (longitude) grid. (Here, r, θ, λ are spherical polar coordinates, t denotes time and B_r the radial component of $\mathbf{B}$.) For each pair we have also calculated the secular change in N , namely

$$\Delta N(t_{11} - t_{10}; r) = N(t_{11}; r) - N(t_{10}; r). \quad (3)$$

The effective vanishing of ΔN at $r = r_c$ provides the theoretical basis of Hide's method for estimating r_c (refs 1, 6).

The computed values of B_r and hence of N and ΔN will depend on n^* , the maximum degree and order of terms of internal origin included in the spherical harmonic expansion, so

Table 1 The r -dependence of N and ΔN

r/r_s	N (10^{12} Wb)				ΔN (10^9 Wb)			
	$n^*=2$, dipole plus quadrupole only							
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
1.00	25.9	25.6	24.8	26.0	-329	-30	+1573	+70
0.95	27.2	27.0	26.2	27.3	-341	-31	+1441	+65
0.90	28.8	28.5	27.7	28.9	-356	-30	+1234	+60
0.85	30.5	30.3	29.4	30.6	-362	-16	+935	+52
0.80	32.5	32.2	31.4	32.5	-353	+7	+490	+42
0.75	34.7	34.4	33.7	34.7	-331	+27	-117	+23
0.70	37.3	37.0	36.5	37.2	-279	+62	-988	0
0.65	40.3	40.1	39.7	40.2	-191			-49
0.60	43.8	43.7	43.7	43.6	-23			-129
0.55	48.1	48.1	48.6	47.7	+275			
0.50	53.4	53.6	55.0	52.7	+825			
	$n^*=3$, dipole, quadrupole and octupole							
	(a)	(b)	(c)		(a)	(b)	(c)	
1.00	25.0	24.8	25.2		+568	+962	+56	
0.95	26.3	26.2	26.8		+431	+1007	-716	
0.90	27.8	27.7	28.7		+158	+888	-1761	
0.85	29.6	29.3	31.0		-339	+559		
0.80	31.7	31.3	33.7		-1200	-122		
0.75	34.2	33.6	37.0			-1200		

See equations (2) (3). Based on the spherical harmonic models (a), (b), (c) and (d) in terms of which Smith, Davis and Jones² analysed Pioneer 10 and Pioneer 11 observations of Jupiter's magnetic field for epochs December 1973 and December 1974.

we have calculated N for $n^*=2$ and $n^*=3$. (We omitted the most highly truncated case of all, $n^*=1$, because both N and ΔN then vary as r_s/r and r_c is indeterminate.) We also investigated the effect of omitting the external contribution to B_r , and found that this made negligible changes to the values of N and ΔN . Table 1 lists the mean value of N and the difference ΔN for each pair of models at $r = r_s(1 - 0.05p)$, with the integer p increasing from zero to at least two steps beyond the point at which ΔN changes sign. The values of N from the various models evidently agree well, even for the smallest values of r , but the values of ΔN are much more varied, and even at $r = r_s$ there is little agreement between the models. Nevertheless, not only do all the models show a change of sign of ΔN for both $n^*=2$ and $n^*=3$, but all of these sign changes occur over a surprisingly small range of r_c , between $0.6r_s$ and $1.0r_s$.

If, by taking $n^*=2$, only the dipole and quadrupole terms are included, the four models give $r_c = 0.60r_s$, $0.82r_s$, $0.76r_s$ and $0.70r_s$ (see Table 1), the mean of which, $0.72r_s$, lies within the range of recent estimates of the radius r_m of Jupiter's metallic core, $0.7r_s$ to $0.8r_s$ (see above). When, in addition to the dipole and quadrupole terms, the less reliable octupole contribution is included (by taking $n^*=3$), only the first three models can be used. These give $r_c = 0.88r_s$, $0.81r_s$ and $0.99r_s$ respectively, mean value $0.89r_s$, corresponding to levels well above the metallic core. (These estimates of r_c should also be compared with the white noise level $0.54r_s$ and the terrestrial analogue level $0.69r_s$, as determined by Elphic and Russell on the basis of considerations of the n -dependence of the quantity

$$(n+1) \left(\frac{r_s}{r} \right)^{2n+4} \sum_{m=0}^n [(g_n^m)^2 + (h_n^m)^2] \quad (4)$$

where g_n^m and h_n^m are the usual spherical harmonic coefficients of order m and degree n ref. 8).)

It is surprising that the new method¹, when applied to models based on data that are so poorly distributed in space and (more importantly) time, should give any value of r_c/r_s at all, let alone values which are both plausible and fairly consistent. While the present results must be regarded as very preliminary, they at least suggest that the method is appropriate and practical for the investigation of the interiors of planets with magnetic fields. It might be possible fairly soon to obtain better estimates of r_c for

Jupiter by combining the Pioneer 10 and Pioneer 11 data with new data from Voyager 1 and Voyager 2, and it might also be possible to use the Pioneer and Voyager measurements of the magnetic field of Saturn to obtain a rough estimate of the radius of the electrically conducting fluid core of that planet.

R. HIDE

*Geophysical Fluid Dynamics Laboratory,
Meteorological Office (Met. O. 21),
Bracknell,
Berkshire, UK*

S. R. C. MALIN

*Geomagnetism Unit,
Institute of Geological Sciences,
Murchison House, West Mains Road,
Edinburgh, UK*

Received 22 March; accepted 23 May 1979.

1. Hide, R. *Nature* **271**, 640-641 (1978).
2. Gehrels, T. (ed.) *Jupiter: Studies of Interior, Atmosphere, Magnetosphere and Satellites* (University of Arizona, Tucson, 1976).
3. Hawke, P. S. *et al. Phys. Rev. Lett.* **41**, 994 (1978).
4. Smoluchowski, R. *Phys. Earth planet. Int.* (in the press).
5. Hide, R. in *Magnetism and the Cosmos* (eds Hindmarsh, W. R., Lowes, F. J., Roberts, P. H. & Runcorn, S. K.) 378-393 (Oliver and Boyd, Edinburgh, 1965).
6. Hide, R. *Geophys. Astrophys. Fluid Dyn.* **12**, 171-176 (1979).
7. Hide, R. & Malin, S. R. C. (in preparation).
8. Elphic, R. C. & Russell, C. T. *Geophys. Res. Lett.* **5**, 211-214 (1978).

A relationship between solar activity and planetary albedos

WE have observed a significant anti-correlation between solar activity and the brightnesses of two Solar System objects. Both the planet Neptune and Saturn's satellite Titan increased in brightness by several per cent between 1972 and 1976¹ and subsequently became fainter by comparable amounts². This period corresponds to the decline of solar activity at the end of solar cycle 20 (1972-76), followed by the rapid increase of activity at the beginning of cycle 21. Solar minimum and the maximum observed brightness of Titan and Neptune both occurred in 1976. (The maximum of cycle 21 is forecast for late 1979 or early 1980, with a sunspot number ~ 150 .) We hypothesise that what we have observed are changes in planetary albedos induced by solar activity. Such changes may have an important bearing on the energy balances of the outer planets and their satellites.

The observed solar-planetary relationship is shown in Fig. 1a for Titan and Fig. 1b for Neptune. Two measures of solar activity were chosen from the several available in the literature: the commonly used Zürich sunspot number, R_z , and an *ad hoc* index, F , the number of observed flares per month of importance 1 or greater. They are shown on the right-hand ordinate scales. The left-hand scales give the brightnesses of Titan and Neptune in the blue (b) and yellow (y) spectral regions in units of stellar magnitudes. The range of brightness variation was ~ 0.08 mag for Titan and 0.03 mag for Neptune, while the range of sunspot number was 0-130.

Figure 1 shows that there are usually two data points per observing season; they represent the mean magnitudes obtained from observations on a number of nights, typically 5-10, preceding and following opposition. The observations are normalised to zero solar phase angle and to mean heliocentric distances of 9.539 and 30.071 AU for Titan and Neptune, respectively. During a single season, each object is measured differentially with respect to two nearby comparison stars whose colour approximates that of the Sun, and subsequent observations determine the relative magnitudes of the comparison stars. A detailed error budget for these observations and a rationale for the method by which they were obtained have been given elsewhere¹.

All the plotted quantities in Fig. 1 are obviously highly correlated with one another. The correlation of the y magnitudes of both objects with the sunspot number is significant at better than the 99% level, while the b magnitude correlations with sunspot number are significant at the 95% level. The correlations of magnitude with the flare index are significant at the ~90% level.

Many directly observed indicators of solar activity (such as 10.7 cm radio flux, coronal line spectrum, number of flares and prominences), as well as solar-terrestrial effects (such as geomagnetic field strength, and cosmic ray flux at the Earth) rise and fall during the 11-yr sunspot cycle, just as the sunspot number does, and would therefore also show a good statistical correlation with the magnitudes of Titan and Neptune. We chose to illustrate the correlation using the sunspot number and flare index.

The observations were made using the 0.5-m reflecting telescope of the Lowell Observatory, a photoelectric photometer equipped with a thermoelectrically cooled EMI 6256 photomultiplier, pulse-counting electronics, and DEC PDP-11 computer. There were no changes in the auxiliary instrumentation during the entire course of the programme. Observations were obtained through two intermediate-band interference filters of the Strömgren photometric system; y at 551 nm (HPBW = 23 nm) and b at 472 nm (HPBW = 20 nm). All

magnitudes are expressed on the natural instrumental scale (that is, not transformed to the 'standard' Strömgren scale) to preserve their inherent high internal accuracy which derived from an extremely stable instrumental configuration.

The change in aspect of a Solar System body as seen from the Earth can cause photometric variations. In the present case these seem to be ruled out. The maximum apparent inclination of Titan's orbital plane (27° with respect to the ecliptic) occurred in May 1974, while maximum brightness was reached in early 1976 when the inclination was 22° . Assuming rough symmetry of the northern and southern hemispheres of Titan (a plausible assumption in view of Titan's thick atmosphere), a seasonal variation would have a period of 15 yr and an amplitude of ~0.2 mag, based on the variations observed by us during one-fifth of a Saturn year. Historically⁴, variations of Titan do not exceed 0.1 mag. Moreover, many configurations of surface markings, if any should be visible through the atmosphere, would lead to substantial variations during Titan's 16-d orbital period. These variations, if they occur at all, are very small (<1%)⁵. The case against seasonal effects on Neptune is even stronger. Since 1972, both the change in aspect as seen from the Earth and the fractional change in solar irradiation were completely negligible.

How the 'solar constant' changes over long periods remains uncertain: the best available Nimbus satellite data suggest that

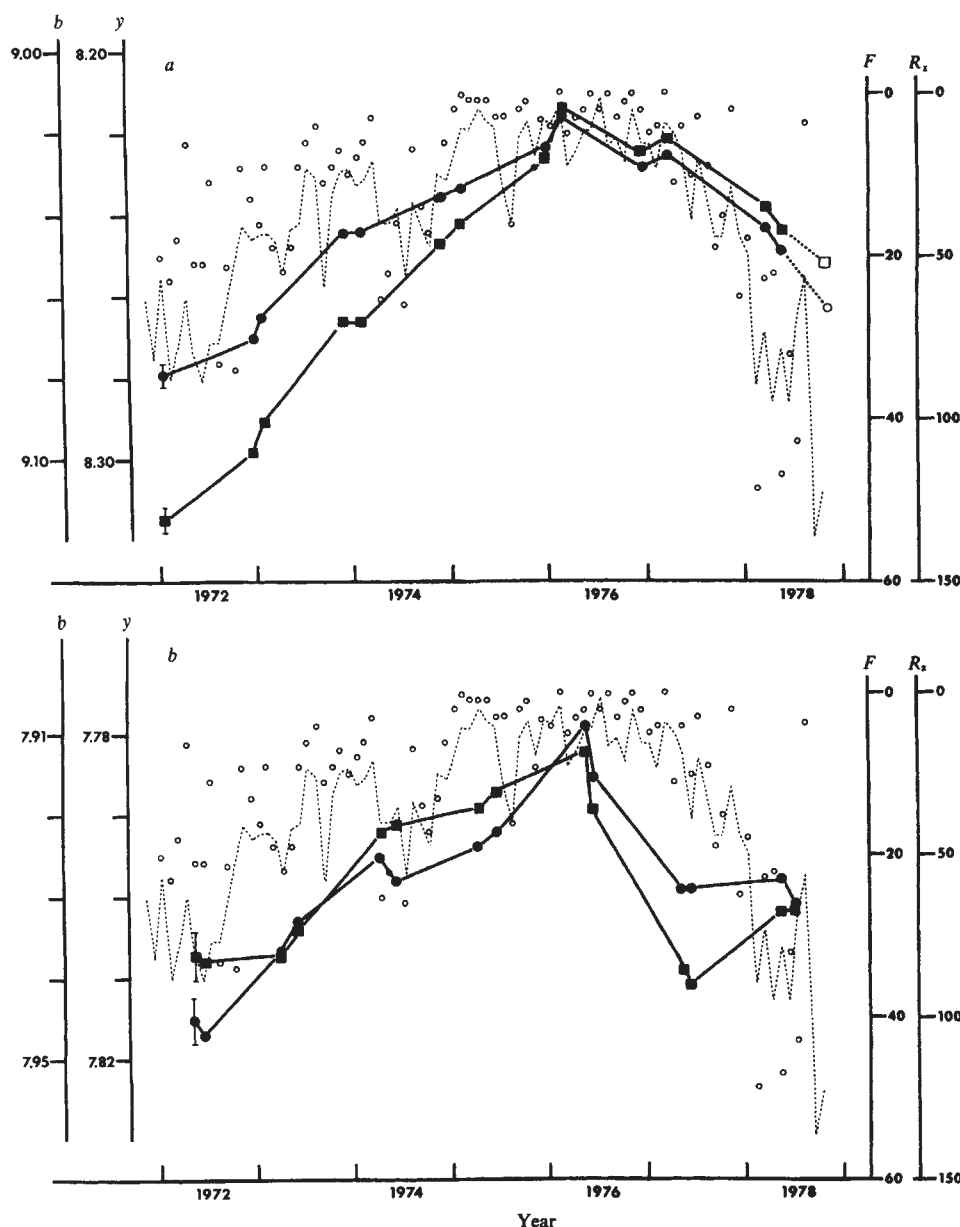


Fig. 1 *a*, Titan and *b*, Neptune. Planetary brightness as a function of time in units of stellar magnitudes (0.01 mag ~ 1%) plotted with brightness increasing upward. Note that the magnitude scales for *a* and *b* are not the same. Typical error bars are shown on the left-most data points. The monthly mean sunspot number R_z (dashed line), and the 'flare index', F ($\circ$), are plotted increasing downward. The solar data are from the *IAU Quarterly Bulletin on Solar Activity and Solar-Geophysical Data*. The flare index was made by counting the number of separate flares per month of importance ≥ 1 tabulated in these publications. $\bullet$, yellow values; $\blacksquare$, blue values.

any changes on a time scale of months are smaller than $\sim 0.2\%$, far too small to account for the effect we have observed.

The apparent synchronisation of solar activity with the observed variations of two separate and independent planetary bodies led us to hypothesise that one or more components of the solar output which vary during the solar cycle could cause changes in the albedos of planets and satellites.

While the thermal structure of Titan's atmosphere remains doubtful, its low albedo in the UV and visible regions implies, according to current models, the existence of a dust layer or photochemical 'smog'. Recent work has explored the possibility of chemical reactions in Titan's atmosphere initiated by cosmic rays, the 'Saturn wind', and the solar wind⁸, as well as photochemical processes leading to the conversion of the abundant CH_4 molecule to heavier hydrocarbons in the atmospheres of the major planets⁹. Such reactions, if they occur, must surely be sensitive to variations of the incident particle or radiative flux; and, therefore, variations in atmospheric composition and structure could also occur, leading to variations in albedo.

Variations of the solar radiative flux below 300 nm during a solar cycle, according to the best available data, range from $\sim 1\%$ at 300 nm to several orders of magnitude below 2 nm^6 . The solar wind and interplanetary magnetic field have recently been sampled in the outer Solar System and have been found to vary in complex ways during the solar cycle^{10,11}.

However, to our knowledge, no studies have been made to consider the consequences of variations of the solar radiative and particle output on planetary atmospheres; and therefore our meagre interpretation of the observed planetary brightness variations is rather speculative. We emphasise that the atmospheric dynamics and chemistry of the major planets and Titan are incredibly complicated and much theoretical work remains to be done before an adequate interpretation will be possible.

In addition to the recent data shown on Fig. 1 we have analysed¹² the more extensive but less accurate Lowell Observatory photoelectric photometry of Uranus and Neptune dating from 1953 to 1966^{13,14} to look for evidence of transient effects such as a planetary response to solar flares, for short-term (month-to-month) correlations with daily and mean monthly sunspot numbers, and for a periodicity in observed brightness at or near the solar rotation frequency. Using correlation techniques, we have looked for time-delayed effects lagging by a few hours (corresponding to propagation at the speed of light), by a few weeks to several months (corresponding to the average speed of the solar wind), and by longer intervals up to one year. None has yet been found at a level of 0.003 mag or greater for Neptune or Uranus, which suggests that transient events may be relatively unimportant.

Thus it seems more likely that it is the long-term envelope of solar activity that drives planetary albedo changes and that these changes derive from relatively slow structural and chemical changes in planetary atmospheres.

This research has been supported since 1972 by the Atmospheric Sciences Program of the NSF.

G. W. LOCKWOOD
D. T. THOMPSON

Lowell Observatory
Flagstaff, Arizona 86002

Received 22 January; accepted 10 April 1979.

- Lockwood, G. W. *Science* **190**, 560–562 (1975); *Icarus* **32**, 413–430 (1977).
- Lockwood, G. W. & Thompson, D. T. *Bull. Am. Astr. Soc.* **10**, 578 (1978).
- Solar Geophysical Data*, No. 409, Part I (Sept, 1978).
- Andersson, L. E. in *Planetary Satellites* (ed. Burns, J. A.), 451 (University of Arizona Press, Tucson, 1977).
- Lockwood, G. W. *Astrophys. J.* **195**, L137–L139 (1975).
- White O. R. (ed.) *The Solar Output and Its Variations* Chs 1, 4 (Colorado Associated University Press, Boulder, 1977).
- Hickey, J. R., Griffen, F. J. & Howell, H. B. *Proc. International Solar Energy Society Solar World Conf.* preprint (1977).
- Chang, S., Scattergood, T., Aronowitz, S., & Flores, J. in *The Saturn System* (eds Hunten, D. M. & Morrison, D.), (NASA Conference Publication 2068, 1978).
- Strobel, D. F. *Rev. Geophys. Space Phys.* **13**, 372–382 (1975).
- Intriligator, D. S. in *Study of Travelling Interplanetary Phenomena 1977* (eds Shea, M. A., Smart, D. F. & Wu, S. T.) 195 ff (Reidel, Dordrecht, 1977).
- Smith, E. J. & Wolfe, J. H. in *Study of Travelling Interplanetary Phenomena 1977* (eds Shea, M. A., Smart, D. F. & Wu, S. T.) 227–ff (Reidel, Dordrecht, 1977).
- Lockwood, G. W. *Icarus* **35**, 79–92 (1978).
- Jerzykiewicz, M. & Serkowski, K. *Lowell Obs. Bull.* **6**, 295–323 (1966).
- Serkowski, K. *Lowell Obs. Bull.* **5**, 157–226 (1961).

Drag reduction in fibre suspensions: transitional behaviour due to fibre degradation

IT is known that the addition of small amounts of certain polymers to a fluid can produce spectacular reductions in turbulent friction^{1,2}. It is also possible to reduce friction by suspending macroscopic fibres in a fluid³. Superficially there are many similarities between the two effects but they are believed to involve different mechanisms because, for example, using polymers and fibres together can produce much larger reductions in drag than would be the case with either additive on its own⁴. We report here measurements of turbulent velocities in a fibre suspension, using laser-Doppler anemometry. We found that the effect of fibres on turbulent structure was not the same as that of polymers. However, when the fibre was degraded by passing it through the system several times, there was a transition to the type of structural change found in polymer solutions.

The fibres used were chrysotile asbestos and are known to be very effective drag reducers. A nominal fibre concentration of 300 p.p.m. by weight was used throughout. Following the usual practice, the fibres were dispersed in 0.5% aqueous solution of Aerosol OT.

A 'blow-down' flow rig was used to minimise degradation of the fibres. Suspensions were prepared in a header tank and then allowed to flow under gravity into a pressure tank. Air pressure was used to propel the suspensions through a horizontal pipe of 2 cm nominal bore and then through a return line to the header tank. Flow rates were controlled by a constant-flow valve. The actual flow rate was determined by timing the change in levels on a sight-glass in the pressure tank. The test section of the pipe was

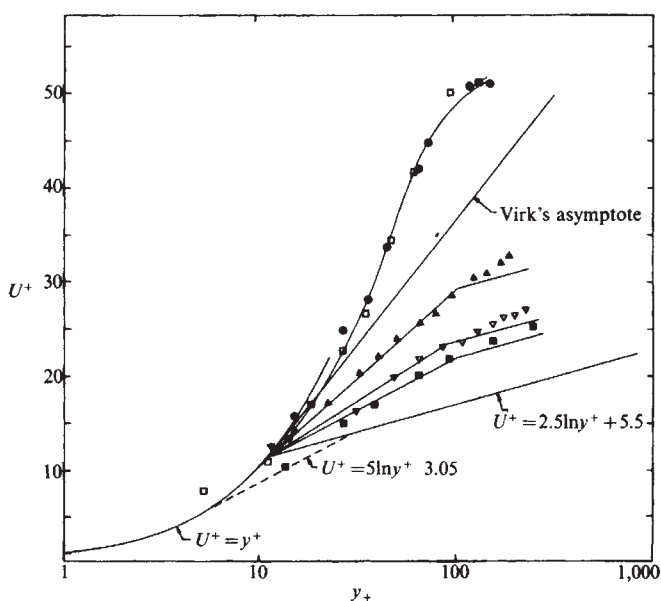


Fig. 1 Mean velocity profiles in a 300 p.p.m. asbestos fibre suspension at various levels of drag reduction for $R_e = 9 \times 10^3$. ● Run 1, drag reduction (DR) = 63%. □ Run 2, DR = 63%. ▲ Run 4, DR = 44%. ▽ Run 10, DR = 22%. ■ Run 17, DR = 17%.

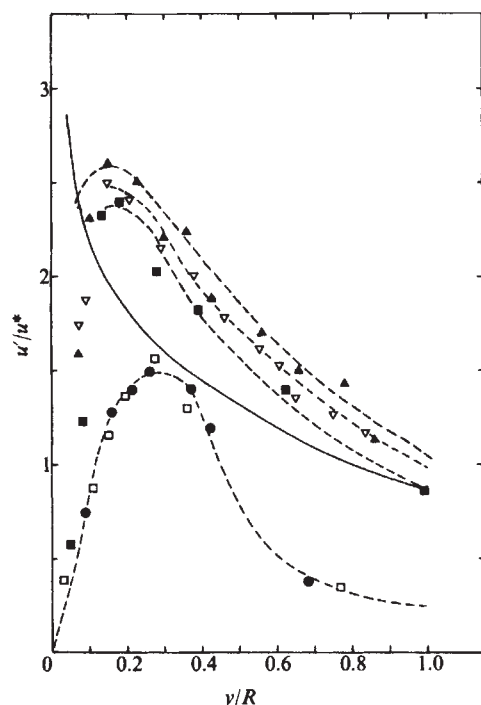


Fig. 2 Streamwise turbulent intensity profiles in a 300 p.p.m. asbestos fibre suspension at various levels of drag reduction, for $Re = 9 \times 10^3$: continuous line denotes results for water. ● Run 1, DR = 63%. □ Run 2, DR = 63%. ▲ Run 4, DR = 44%. ▽ Run 10, DR = 22%. ■ Run 17, DR = 17%.

situated at 154 diameters from the inlet to ensure well developed flow. Pressure drops were measured using pressure taps and a liquid-in-glass manometer.

Streamwise velocity profiles were measured over the cross-section of the pipe, using a laser-Doppler anemometer (LDA). The optical and signal-processing arrangements were as reported previously⁵.

An experimental 'run' consisted of one pass of the whole amount of working fluid through the apparatus at a fixed Reynolds number. Further 'runs' were further passes of the same fluid at the same Reynolds number. Experimental results were always taken in steady conditions, avoiding the beginning and end of each run. Reynolds numbers were calculated from the pipe diameter, the bulk mean velocity and the kinematic viscosity of water.

These particular fibres are known to be susceptible to degradation under shear. Our results confirmed this and showed a decline in drag-reducing effectiveness with increasing run number. We decided to make a virtue of necessity and used this as an experimental parameter. The results were interesting.

Figure 1 shows mean velocity U^+ (scaled) plotted against distance from the pipe wall y^+ (scaled), in the usual 'wall-region' coordinates, for a Reynolds number of 9×10^3 . The amount of drag reduction appears as a parameter. It can be seen that the first two runs gave a drag reduction of 63% and that the corresponding mean velocity profiles lay beyond the Virk asymptote for polymer solutions. In contrast, the results for later runs in the degraded suspensions were in the polymeric regime. Like results for polymers, they show a trend to the newtonian result, as the amount of drag reduction decreases.

Figure 2 shows the r.m.s. velocity u' (divided by the friction velocity u^*) plotted against distance from the wall y (divided by the pipe radius R). Here the differences between high and low drag reduction are seen to be no longer a matter of degree only. Runs 1 and 2 showed a marked suppression of the streamwise fluctuating velocity, compared with the newtonian result. However, in later runs the intensity over most of the cross-section was increased above the newtonian result. Again, these

results in later runs were similar to those previously found in polymer solutions¹.

This behaviour has also been found in further experiments at a higher Reynolds number, 1.4×10^4 . Thus it seems that we are observing a transition between two different drag-reduction mechanisms. When the fibre is at maximum effectiveness, it behaves quite differently from polymers which produce comparable amounts of drag reduction. In particular, it suppresses the streamwise turbulence. After degradation, when the fibre is (presumably) smaller it seems to behave rather like a polymer of the same effectiveness. In particular, it enhances the streamwise turbulence.

It might be tempting to conclude from Figs 1 and 2 that the undegraded fibres reduce drag by suppressing turbulence and making the flow 'more laminar'. Such a conclusion would be wrong. In our more recent experiments we have used a second LDA to measure the spanwise turbulent intensity. During runs 1 and 2, the spanwise turbulent intensity was actually larger than in water alone. In the later runs, the spanwise intensity was less than that for water alone. Rudd has previously reported⁶ reduced spanwise intensities in polymer solutions.

We already know that the turbulent structure is changed during drag reduction in polymer solutions^{1,6}. Our present results indicate that in undegraded fibre suspensions the change in turbulent structure during drag reduction is of a different kind. This effect was previously predicted on theoretical grounds⁷.

We thank Turner Brothers Asbestos Co. Ltd for drag-reducing chrysotile asbestos suspensions.

W. D. MCCOMB
K. T. J. CHAN

School of Engineering,
University of Edinburgh,
Edinburgh, UK

Received 23 April; accepted 30 April 1979.

1. Virk, P. S. *A.I.Ch.E. J.* **21**, 625 (1975).
2. Hoyt, J. W. *Proc. 2nd Int. Conf. on Drag Reduction*, Paper A1, Cambridge, UK (BHRA, 1977).
3. Hoyt, J. W. *Naval Undersea Center Rep. TP299*, San Diego (1972).
4. Lee, W. K., Vaseleski, R. C. & Metzner, A. B. *A.I.Ch.E. J.* **20**, 128 (1974).
5. McComb, W. D., Allan, J. & Greated, C. A. *Phys. Fluids* **20**, 873 (1977).
6. Rudd, M. J. *J. Fluid Mech.* **51**, 673 (1972).
7. McComb, W. D. *Nature phys. Sci.* **241**, 117 (1973).

Ferromagnetism in ancient copper-based coinage

WHEN extensive forgery of silver coinage has taken place in the past, a small permanent magnet was a well known technique for distinguishing between counterfeit nickel and genuine silver coins. The accidental discovery of ferromagnetism in undisputedly genuine ancient Kushan copper coins by Nisbet during a routine check with a permanent magnet was, however, totally unexpected. Although we have found very few coins possessed a magnetic moment sufficiently large for the coin to be lifted by the permanent magnet, subsequent examination has revealed that several coins possess a spontaneous remanence magnetisation which can be detected with a simple deflection magnetometer. Quantitative measurements of the room temperature magnetisation have been performed in fields up to $5 \times 10^5 \text{ A m}^{-1}$ using a Faraday balance magnetometer. To our knowledge they form the first systematic study of the magnetic properties of ancient coinage.

The magnetometer sensitivity was about $10^{-2} \text{ J T}^{-1} \text{ kg}^{-1}$ and the Pauli paramagnetism of the conduction electrons could not be detected. As Fig. 1 shows, the magnetisation of an orichalcum semis of the Roman Emperor Tiberius remained below the noise level up to the maximum fields available. Similar null results

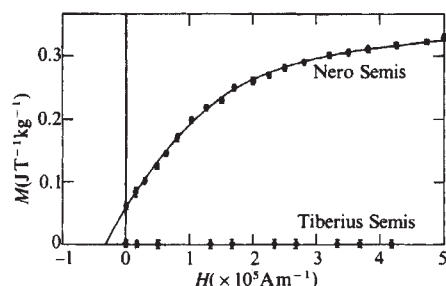


Fig. 1 Magnetisation as a function of field for two Roman orichalcum coins. The magnetisation of the Tiberius semis was at all times well below the instrumental sensitivity.

were obtained with modern bronze coinage, entirely consistent with the expected paramagnetic behaviour. In marked contrast, an orichalcum semis of the Emperor Nero (dated to 63–64 AD) not only showed significant remanence magnetisation but also a marked increase with increasing magnetising field. Although the susceptibility fell as the field was raised, no tendency towards saturation was observed. When the field was reduced, the remanence magnetisation returned to its initial value. The coercive field H_c (that is, the reverse field required to reduce the magnetisation to zero) was found to be $3.2 \times 10^4 \text{ A m}^{-1}$ (400 Oe) and the remanence magnetisation M_r to be $5.5 \times 10^{-2} \text{ J T}^{-1} \text{ kg}^{-1}$.

Orichalcum, first analysed in 1798 by Klaproth¹, is an alloy of copper and zinc and is generally believed to have been produced by a cementation process, in which copper bars were heated in a crucible containing a zinc ore and charcoal². Such an alloy should not possess a spontaneous magnetic moment. Recent analyses have shown, however, that some Roman orichalcum coins may contain up to 1% iron impurity³, and with this in mind magnetisation measurements were performed on a series of coins exhibiting ferromagnetism which had been previously analysed by Thompson using spark emission spectroscopy.

Although neglected in secondary school teaching in the West, the Kushan Empire was the third of the great empires of antiquity controlling the caravan routes between Rome and Cathay. At its height it covered what is now Afghanistan, Pakistan, North-west India and Southern Russia.

Magnetisation measurements were made on 12 copper Kushan coins of known chemical composition. A typical magnetisation curve is shown in Fig. 2a. All coins showed ferromagnetic behaviour, and saturated in fields of about $1.5 \times 10^5 \text{ A m}^{-1}$ in marked contrast to the Roman orichalcum coin. The coercive field was low, having a mean value of $3.9 \times 10^3 \text{ A m}^{-1}$ (with a 1σ error of $1.2 \times 10^3 \text{ A m}^{-1}$). Although both the shapes of the magnetisation curves and the values of the

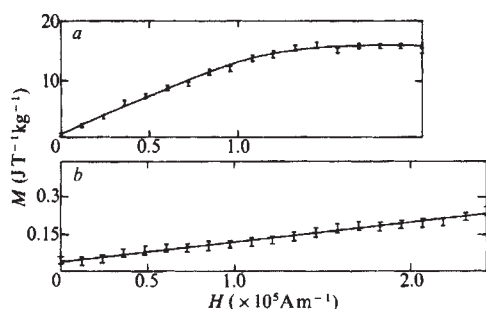


Fig. 2 a, Magnetisation curve for a copper drachm of Kanishka (Kushan Empire). Note the saturation, small remanence magnetisation and low coercive field. **b**, Magnetisation curve for a similar Kashmiri coin. Note the high coercive field, and lack of saturation.

coercive fields were all very similar, two distinct groups of coins were identified. Coins of the late Kushan period (mid-third to mid-fourth century AD) displayed a very low saturation magnetisation M_s and remanence so small as to be undetectable with a deflection magnetometer. Coins produced during the period of great expansion of the Kushan Empire under Kanishka (second quarter of second century AD) showed saturation and remanence magnetisation two to three orders of magnitude greater than those of the late Kushan period. These remanence magnetisations were easily measured with a deflection magnetometer. The strong ferromagnetism was thus confined to coins of a restricted historical period.

In contrast, early Hindu copper coins from Kashmir, showed markedly different magnetisation curves. Nine coins whose dates ranged over a wide period were examined. All showed ferromagnetic behaviour but, as with the orichalcum coins, no saturation occurred up to the highest fields available (Fig. 2c). A large spread occurred in the coercive fields but all were significantly larger than that of the Kushan coinage. The mean value of $2.9 \times 10^4 \text{ A m}^{-1}$ (with 1σ error of $1.8 \times 10^4 \text{ A m}^{-1}$) is nearly an order of magnitude greater than the mean value for

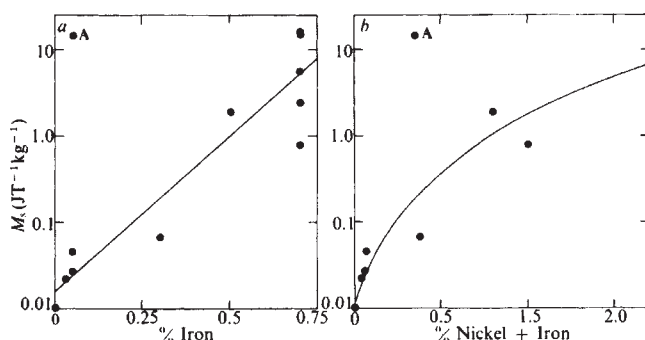


Fig. 3 a, Log-linear plot of saturation magnetisation versus iron content in Kushan coinage. The straight line is a least squares fit to the data. The point A was rejected in the fit on the grounds of unreliability of the analysis. **b**, Log-linear plot of M_s versus total iron and nickel content. The line is a guide for the eye.

Kushan coins and is comparable with that of the Roman orichalcum. No correlation has yet been established between strong ferromagnetism in the Kashmiri coins and the historical period.

A correlation does exist between concentration of iron and the saturation magnetisation M_s in the Kushan coinage. The chemical analyses show the lead, nickel and iron content of the Kanishka coins to be of the order of 1%, significantly above that of the late Kushan coins. Nickel, though ferromagnetic itself, forms a complete solid solution in copper and a 99% Cu–1% Ni alloy is not ferromagnetic at room temperature. However, iron is almost insoluble in copper and a dispersed-phase alloy of 99% Cu–1% Fe would have a magnetisation comparable with the largest values measured. Figure 3a shows a log-linear plot of M_s versus iron content. Clearly, the saturation magnetisation is directly related to the iron content. Fitting a straight line to the data by the least squares technique yielded a correlation coefficient of 0.94. However, a correlation still persisted when the total ferromagnetic impurity concentration (Fe + Ni) was plotted against M_s (Fig. 3b). As iron and nickel form a range of ferromagnetic alloys it is unclear what role the nickel plays and until more information on the structure of the material is available care should be exercised in interpreting these data. As yet it is unknown whether the iron is present as the element or as an oxide.

The ratio of saturation to remanence magnetisation (M_s/M_r) versus iron content is shown in Fig. 4. M_s/M_r is approximately the same for all the coins and there is no significant dependence on iron (or nickel) content.

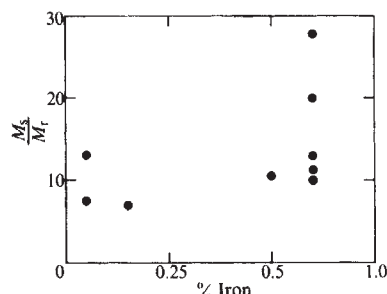


Fig. 4 Ratio of saturation to remanence magnetisation versus iron content in Kushan coinage. (Note that this is a linear plot, unlike Fig. 3.)

We may summarise the results of these initial investigations as follows. (1) Ferromagnetism in ancient coinage arises from the presence of ferromagnetic impurity in the copper-based matrix. (2) The occurrence of ferromagnetism may well have archaeological significance. Very high values of saturation magnetisation in the Kushan coinage are confined to a relatively short period during which time the empire was rapidly expanding. There is some numismatic and historical evidence to suggest that these coins may have been minted from freshly mined copper. (3) The magnetic characteristics (for example, the coercive field) are very different in the otherwise similar Kushan and Kashmiri coins. The magnetic properties may well give a clue to the origin of the ore. High coercive field and high saturation field could be strong indicators of re-usage of old coinage in the minting process. (4) Within a factor of three, the ratio of saturation to remanence magnetisation of the Kushan coinage is constant. Thus M_r can be used to give an order of magnitude estimate of M_s . As M_r can be measured with a simple deflection magnetometer, useful physical measurements can be performed on confined collections such as that in the British Museum.

We thank J. D. F. Nisbet for information about his accidental discovery, and Professor F. C. Thompson, for performing the chemical analyses on the coins studied. I.B.M. acknowledges a maintenance grant from the SRC.

B. K. TANNER
D. W. MACDOWALL*
I. B. MACCORMACK
R. L. SMITH

Department of Physics,
University of Durham,
South Road, Durham, UK

Received 27 April; accepted 23 May 1979.

Present addresses: *Polytechnic of North London, †AERE, Harwell, Didcot, Oxon, UK.

1. Klaproth, M. H. *Allgm. J. Chem.* **6**, 227 (1801).
2. Forbes, R. J. *Metalurgy in Antiquity*, 277 (Leiden Z 1950).
3. Caley, E. R. *Orichalcum and Related Ancient Alloys* (Numismatic Notes and Monographs No. 151, American Numismatic Society, New York, 1964).

Confirmation of the Suess wiggles: 3200–3700 BC

USING dendrochronologically dated tree rings de Vries¹ showed that the atmospheric ^{14}C concentration has not been entirely constant during the past few hundred years. By extending the North American bristlecone pine tree-ring chronology² to nearly 8,000 yr ago, several radiocarbon laboratories succeeded in measuring deviations in the ^{14}C concentration from the mid-nineteenth century natural ^{14}C level to a maximum of about 10% around 4000 BC (refs 3–7). Superimposed on this general trend, Suess drew with 'cosmic schwung', medium-term variations through his data. These so-called 'wiggles' should

correspond with variations in the abundance of radiocarbon in atmospheric carbon dioxide on a time scale between one decade and a few centuries. The resulting irregular shape of Suess' calibration curve has been questioned during the past few years^{8,9}. The overall result, however, is that no generally accepted ^{14}C calibration curve is yet available. It would be most convenient to measure ^{14}C concentrations in single tree rings with much higher precision than the usual 4–6%. For this, however, much larger samples are required than are available from the bristlecone pine tree. An excellent opportunity was presented by the sub-fossil oak trunks discovered in the river valley sediments of southern Germany⁹; our measurements of these trunks are reported here.

Our first measurements were carried out on a 500-yr section of the 1,235-yr floating South German Neolithic tree-ring chronology. The section used was derived from well preserved sub-fossil oaks, which grew along the River Danube (49°25'N, 10°5'E). To avoid recycling due to soil CO_2 during an early stage of tree growth no rings from near the centre of a log were analysed. Because recent results¹¹ have shown that if year to year ^{14}C variations are small then radial movement of ^{14}C across ring boundaries in an oak is negligible, the measured ^{14}C concentration in a single tree ring nicely represents the atmospheric ^{14}C concentration in the growth season of a tree ring. Although wood, after pretreatment with successive solutions of 4% hydrochloric acid, 4% sodium hydroxide and 4% hydrochloric acid again (AAA) probably do not exactly originate from exactly the same time of the year as the cellulose fraction, the AAA pretreatment seems to be acceptable for monitoring past ^{14}C levels¹¹. The final sample prepared for analysis was 50–55 l of CO_2 gas at STP. The special high-precision proportional gas counting system will be described elsewhere¹².

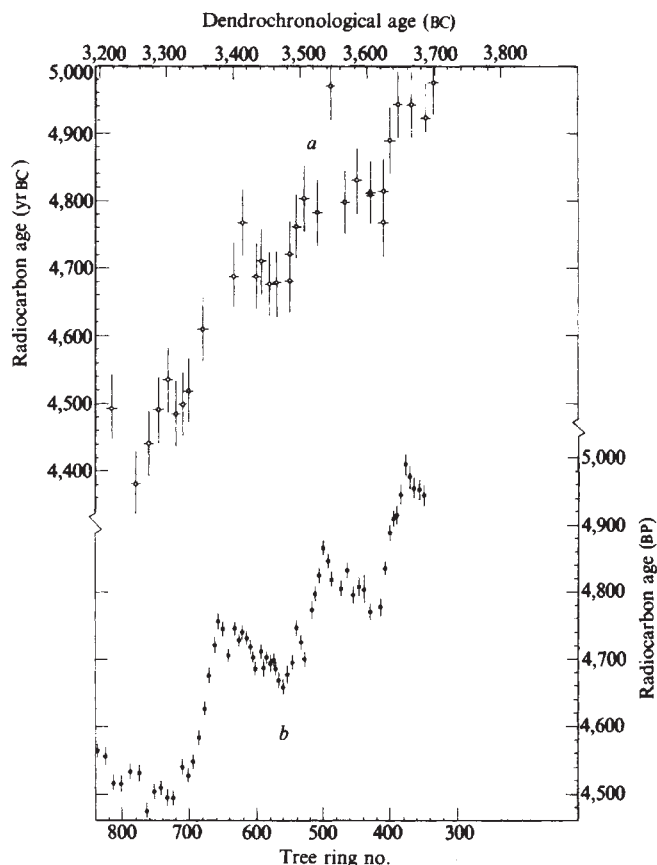


Fig. 1 ^{14}C analyses on dendrochronologically dated wood (b). The relation between radiocarbon age and the tree-ring number from the floating South German Neolithic master chronology. For comparison, a shows the bristlecone dates, corrected to the proper standard activity ($0.95 \times \text{NBS-oxalic acid}$)⁶.

The results are presented as conventional ^{14}C ages in Fig. 1b, related to the tree-ring number from the floating South German Neolithic master chronology.

Figure 1 confirms quantitatively the existence of pronounced wiggles between the thirty-third and thirty-seventh centuries BC. The fact that the upward slopes from right to left are less steep than the downward slopes suggests that we are faced with secular variations in the atmospheric ^{14}C level consisting of relatively fast rises followed by a restoration of the stationary state. This conclusion is supported by the rather regular periodic behaviour of the 'wiggles' with a period of approximately 180 yr. Calculations based on geophysical and geochemical models might provide a more definite answer to the question of the origin of the variations.

A comparison with the bristlecone dates of Suess⁶ corrected to the proper standard activity (0.95 NBS oxalic acid), which makes the dates approximately 50 yr older than listed, leads to the absolute time scale for our floating chronology (Fig. 1a). The South German Neolithic master chronology seems to start at 4035 ± 3 BC. Wiggle matching Suess' results on the German material with the bristlecone sequence gave a comparable result (4057 BC)^{10,13}.

For archaeological samples which are 'short lived' (of the order of a few tens of years) these medium-term variations add an extra uncertainty to the purely statistical error of a ^{14}C -age determination.

A. F. M. DE JONG
W. G. MOOK

Isotope Physics Laboratory,
University of Groningen,
The Netherlands

B. BECKER

Botanisches Institut,
Universität Stuttgart-Hohenheim FRG

1. de Vries, H. L. *Kon. Ned. Akad. Wet. Proc.* B61, 94–102 (1958).
2. Ferguson, C. W. *Proc. 8th int. Conf. Radiocarbon Dating*, A1-11 New Zealand (1972).
3. Damon, P. E., Long, A. & Wallick, E. I. *Proc. 8th int. Conf. Radiocarbon Dating*, A28–44 (1972).
4. Ralph, E. K. & Michael, H. N. *12th Nobel Symp.*, 237–245 (1970).
5. Suess, H. E. *Radiocarbon Dating and Methods of Low Level Counting*, 143–151 (IAEA Vienna, 1967).
6. Suess, H. E. *Radiocarbon* 20, 1–17 (1978).
7. Suess, H. E. *Proc. 9th int. Radiocarbon Conf. UCLA and UCSD* (in the press).
8. Clark, R. M. & Renfrew, C. *Archaeometry* 14, 5–19 (1971).
9. Pearson, G. W., Pilcher, J. R., Baillie, M. G. L. & Hillam, J. *Nature* 270, 25–28 (1977).
10. Becker, P. *Proc. IXth int. Radiocarbon Conf. UCLA and UCSD* (in the press).
11. Tans, P. P., de Jong, A. F. M. & Mook, W. G. *Nature* 271, 234–235 (1978).
12. Tans, P. P. & Mook, W. G. *Radiocarbon* (in the press).
13. Suess, H. E. & Becker, B. *Proc. Symp. Die Dendrochronologie des Postglazials*, 156–170 (Akad. d. Wiss. Mainz 1977).

Sulphur uptake from the atmosphere by forest and farmland

MEASUREMENTS of sulphur uptake by plants are needed to aid our understanding of the long-range transport of sulphur dioxide and the role of sulphur uptake in acidification and the nutrient balance of forest ecosystems¹. The traditional profile method of analysing sulphur uptake has proved inadequate for forest measurements because the vertical concentration gradients above forests are small and difficult to measure with current techniques²; the best estimates available are based on extrapolation of measurements from individual shoots to the whole forest³. We report here the first direct measurements of the uptake of sulphur from the atmosphere by a forest in the absence of precipitation (that is, dry deposition). These measurements were possible because of the development of a fast response sulphur detector which allowed, for the first time, use of the eddy correlation method⁴ for sulphur uptake measurements. The sulphur uptake ranged from 0 – $0.1 \mu\text{g m}^{-2} \text{ s}^{-1}$ (as SO_2) for atmospheric sulphur concentrations up to $35 \mu\text{g m}^{-3}$ (expressed as SO_2). On average, $\sim 75\%$ of the

atmospheric sulphur was in gaseous form. These uptake rates support those derived from use of the recent hypothesis that sulphur dioxide uptake by vegetation takes place through the stomatal opening for a dry canopy³. Some measurements of dry uptake of sulphur over farmland were also made and are in close agreement with published measurements made over similar surfaces using the profile and other methods.

The eddy correlation method can give a direct measurement of the vertical flux of a constituent within the air. If the concentration of the constituent is fairly constant (that is, storage is negligible), the vertical flux within the atmosphere is equal to the uptake rate at the underlying surface. At a fixed point above the surface, continuous fast response measurements are made of the vertical wind velocity, w , and the concentration (ambient density) of the constituent, c . The vertical flux of the constituent, F , is given by the covariance $F = \overline{wc}$ where the overbar indicates the time average. We assume that the time averaged vertical velocity is zero, $\bar{w} = 0$, (ref. 5), therefore $F = \overline{w'c'}$ where the prime indicates the instantaneous departure from the long-term mean. This uptake of the constituent from the atmosphere is conveniently expressed in terms of a deposition velocity, v , and the air concentration of the constituent, c , by an equation $F = vc$ (refs 1–3, 6).

A Gill propeller anemometer was used to measure the vertical velocity. While the response time τ of a Gill propeller (that time to reach $1 - 1/e$ of the final speed following a step change in wind speed) depends on mean wind speed U , a response length $L = U\tau$ should depend only on the turbulent intensity and L was estimated to be 2 and 3 m for the Crowthorne (forest) and Blewbury (farmland) measurements respectively^{7,8}. The flame photometric sulphur detector⁹, with a 4-m gas inlet tube, had a response time measured in the field of 0.3 s. This response time was determined mainly by the inlet tube. There was also a time delay of 1.8 s in the gas-inlet system of the sulphur detector. This was corrected for in the covariance calculation by displacing digital records from the sensors by this time interval. Ambient air temperature was measured using a fast-response bead thermistor to determine the sensible heat flux. The sensors were sampled at 5 Hz by a digital computer for the covariance calculations. No low frequency filter was used in the system but fluxes were computed for individual 10- or 12-min periods.

The flux values were corrected^{7,8} due to inadequate speed of response of the anemometer and to smoothing from the sulphur detector inlet system to allow for flux losses. The corrections were 10 and 20% for the Crowthorne and Blewbury data respectively. Low frequency flux losses, probably of the order of 5–20% (ref. 10) have not been corrected for in the flux estimates. The ratio of height of sampling to length of uniform fetch upwind of the sensors was such that even in the worst case the flux should be 90%, adjusted to the underlying forest surface. Adjustment was often complete¹¹.

Measured non-zero values of $\bar{w}$ indicated slight misalignment of the vertical propeller amounting to $\sim 1^\circ$. This misalignment introduces an error into the mass flux measurement of the order of 4% per degree^{5,12}. Considering these factors we estimate the absolute accuracy of the average flux measurements to be of the order of $\pm 50\%$. The short sample period will cause a scatter in individual flux values of $\pm 25\%$ (refs 13, 14).

Measurements in the field suggest that the standard deviation in a single 10-min flux measurement due to instrument noise alone would be ± 0.02 to $\pm 0.03 \mu\text{g m}^{-2} \text{ s}^{-1}$ and that the standard error in average values from hours of observations as in Table 1 from instrument noise alone would be perhaps $\pm 0.006 \mu\text{g m}^{-2} \text{ s}^{-1}$. A description of the full system along with a complete analysis of the data will be published elsewhere.

Measurements were made at two sites (see legend to Table 1). At the farmland site the equipment was located at the boundary between two fields. In an easterly wind the fetch was entirely over a grass crop, and in a westerly wind, over a pea crop. Table 1 shows the average uptake rate for each range of concentration and is composed from all the individual flux measurements made in the concentration range. The data suggest a deposition

Table 1 Sulphur uptake measurements

Site and times of measurements	Period of measurement (h)	Air concentration ($\mu\text{g m}^{-3}$)	Gaseous fraction (%)	Mean uptake $\pm$ s.e.m. ($\mu\text{g m}^{-2} \text{s}^{-1}$)	Wind (m s^{-1})
Forest*					
1030–2000 h	16	6–11	60	<0.01	3.2–4.8
1200–1700 h	7	12–35	84	0.04 ± 0.02	0.8–4.2
Farmland†					
1200–1730 h	3	3–8	50	0.03 ± 0.02	2.2–5.7 (westerly; peas)
1100–1600 h	5	16–50	75	0.19 ± 0.02	2.2–4.4 (easterly; rye grass)

* Forest: Crowthorne, Berkshire. 30-yr-old Scots Pine 14 m high unthinned since planting. Flux measurement 3 m above tops. Measurements are for 9 days between 1 and 20 May 1977.

† Farmland: Blewbury, Berkshire. Easterly fetch; Italian rye grass 75 cm; westerly fetch, peas 40 cm. Flux measurement 4 m above soil surface. Measurements are for six days between 21 June and 5 July 1977.

velocity of about 0.2 cm s^{-1} for these forest measurements and about 0.6 cm s^{-1} for the farmland measurements. All measurements were made in a turbulent atmosphere.

The sulphur detector responds to gaseous and particulate compounds. Particulate sulphur contributes only a fraction of the fluxes reported here, as the sulphate concentration at the field sites, assumed to be similar to that measured at Harwell, was only 20% of the total sulphur at the higher concentrations and the deposition velocity for small particles such as sulphate is much smaller than, or comparable with, that for sulphur dioxide^{2,15}

Garland² summarises measurements of sulphur dioxide deposition over grass, crops and bare soil made by the profile and radioactive methods that give deposition velocities in the range $0.3\text{--}1.2 \text{ cm s}^{-1}$ with a median value 0.8 cm s^{-1} , in close agreement with that measured by the eddy correlation method here. Profile measurements in a forest² suggested an upper limit of the deposition velocity of 2 cm s^{-1} . Estimates based on measurements of individual shoots and extrapolation suggest a deposition velocity varying from a daytime maximum value of 0.6 cm s^{-1} in the morning to 0.2 cm s^{-1} in late afternoon to a night-time value of 0.1 cm s^{-1} , depending on stomatal resistance (closure) for a dry forest canopy³. The measurements presented here for the higher sulphur concentrations suggest a similar low deposition velocity for sulphur on the forest canopy during the afternoon. (We found no evidence of a diurnal variation in surface resistance during the period of our measurements.) Forest measurements were made when the new growth on the forest was just emerging and most of the exposed needles would have been from previous year's growth and are known to display lower uptake rates³. Furthermore, the sensible heat flux from the forest was frequently about 50% of the global total incoming solar radiation. This indicates that little evapotranspiration was taking place. The stomatal resistance to gaseous transfer would most probably be high in these conditions. Indeed, stomatal resistance is generally higher during the afternoon^{3,16}, and the measurements presented here may be for a forest in conditions of maximum resistance to sulphur uptake. The observed fluxes over forest and farmland can be entirely accounted for in terms of sulphur dioxide deposition. However this does not preclude a large deposition velocity for particulate sulphur.

The dry deposition to a dry forest canopy at a mean sulphur dioxide concentration of $10 \mu\text{g m}^{-3}$, typical of many rural locations, would be perhaps $2\text{--}3 \text{ kg sulphur ha}^{-1} \text{ yr}^{-1}$, if the representative deposition velocity is as low as that measured here. This is a significant fraction of the estimated total sulphur deposition in English and Scottish upland sites of $10\text{--}20 \text{ kg sulphur ha}^{-1} \text{ yr}^{-1}$ (ref. 1). If as suggested³ gaseous and particulate uptake onto a wet canopy could account for a much larger fraction of the total deposition, then measurements with the method presented here can resolve this question.

We thank L. C. Cox for help with the field work, A. Hadjitofi for developing the sulphur detector, M. M. Davies for pro-

gramming the computer, C. Healy and Miss J. B. Tarrant for the sulphate analyses, the Director of the Road Research Laboratory and the Manager of Churn Estates farm for access to field sites. I.E.G. thanks the Nuffield Foundation for a Travelling Fellowship and Dr A. C. Chamberlain for his help during the visit. The work was supported by the Department of the Environment.

I. E. GALBALLY

CSIRO Division of Atmospheric Physics,
P.O. Box 77, Mordialloc,
Victoria, Australia 3195

J. A. GARLAND

Environmental and Medical Sciences Division,
AERE, Harwell,
Oxfordshire, UK

M. J. G. WILSON

Department of Chemical Engineering and Chemical Technology,
Imperial College, Prince Consort Road,
London SW7, UK

Received 27 December 1978; accepted 15 May 1979.

1. Department of Environment, Pollution Paper No. 7, *Effects of Airborne Sulphur Compounds on Forests and Freshwater* (HMSO, 1976).
2. Garland, J. A. *Proc. R. Soc. A* **354**, 245 (1977).
3. Garland, J. A. & Branson, J. R. *Tellus* **29**, 445 (1977).
4. Priestley, C. H. B. *Turbulent Transfer in the Lower Atmosphere* (University of Chicago Press, 1959).
5. Desjardins, R. L. & Lemon, E. R. *Boundary-Layer Met.* **5**, 475 (1974).
6. Galbally, I. E. *Adv. Geophys.* **18B**, 329 (1974).
7. Hicks, B. B. *Boundary-Layer Met.* **3**, 214 (1972).
8. Garratt, J. R. *Boundary-Layer Met.* **8**, 255 (1975).
9. Hadjitofi, A. & Wilson, M. J. G. *Atmos. Envir.* (in the press).
10. McBean, G. A. *J. appl. Met.* **11**, 1078 (1972).
11. Dyer, A. J. & Garratt, J. R. *J. met. Soc., Japan* **56**, 19 (1978).
12. Dyer, A. J. & Hicks, B. B. *Q. J. R. met. Soc.* **98**, 206 (1972).
13. Wyngaard, J. C. in *Workshop on Micrometeorology* (ed. Haugen, D. A.) (Am. Met. Soc., 1973).
14. Sreenivasan, K. R., Chambers, A. J. & Antonia, R. A. *Boundary-Layer Met.* **14**, 341 (1978).
15. Wesely, M. L., Hicks, B. B., Dannevik, W. P., Frisella, S. & Husar, R. B. *Atmos. Envir.* **11**, 561 (1977).
16. Gash, J. H. C. & Stewart, J. B. *Boundary-Layer Met.* **8**, 453 (1975).

Magnetic measurements used to assess sediment influx at Llyn Goddionduon

THE preliminary results of an attempt to estimate sediment influx in a small upland lake using magnetic susceptibility measurements on whole unextruded sediment cores obtained from a grid covering almost the whole area of open water are reported here. The susceptibility variations, reinforced by a wider range of magnetic measurements on individual samples, provide a basis for core correlation over the whole lake. The speed with which such studies can be accomplished makes possible a much more secure empirical basis for calculations of total sediment input versus time than do conventional methods of core correlation and analysis.

Increasing attention is being devoted to using lake sediments as sources of information about changing ecological processes in the recent past^{1,2}. In many cases a large proportion of the lake sediment can be shown to have been derived from the lake watershed. The palaeoenvironmental record may, therefore, often be interpreted largely in terms of changes in the terrestrial ecosystems draining into the lake^{3,4}. A prerequisite of fully quantitative studies designed to relate the sedimentary record in the lake to changing material output from its catchment is some procedure for calculating or estimating sediment influx to the whole lake for defined time intervals. Without this procedure all calculations of influx may be biased by the unrepresentative nature of a small number of coring sites.

Empirical attempts to overcome this problem⁵ are limited largely by the time-consuming nature of the techniques normally available for core correlation in the absence of synchronous visual stratigraphic markers. Theoretical models developed to allow calculation of total sediment input from a single central core⁶, though valuable are unconvincing in view of the numerous unverified assumptions on which they depend.

Thompson *et al.*⁷ have shown that magnetic susceptibility measurements of whole, unextruded sediment cores can provide a rapid method of core correlation. The technique leaves the material unaltered, precludes no subsequent study of any kind and can be followed up by a range of related simple, rapid and non-destructive magnetic measurements⁸ which often indicate

sediment source^{9,10} and are directly related to changes in human activity^{7,8,11}.

Llyn Goddionduon (Fig. 1a) lies 244 m above OD in the Gwydyr Forest area of North Wales, ~4 km north-west of Bettws y Coed (G.R. SH 753 586). It covers 6.2 hectares and receives drainage from a catchment of ~25 hectares. Maximum water depth is 6 m, in a small 'trench' occupying <5% of the area of the lake bed. Of the present sediment surface, 50% is between 0.75 m and 4 m deep. A small mire ~0.6 hectares in extent at the southern, outflow, end of the lake represents overgrowth of peat lying partly on marginal lake sediments. Some 25% of the lake is colonised by macrophytes among which *Schoenoplectus lacustris*, *Phragmites communis*, *Eleocharis palustris*, *Nymphaea alba*, *Potamogeton natans* and *Equisetum fluviatile* are the most abundant species¹².

An accurately surveyed grid of 130 minicores¹³ was sampled during a 2-week period in July 1977, using a research team varying from five to nine people. Cores were taken at 20 m × 20 m intersections. Where impenetrable substrate precluded this, intermediate cores were taken. In the small trench area an extremely high density of sampling was achieved (27 cores in 950 m²). Whole core susceptibility measurements were carried out on all cores during the same period of field work using a low-field susceptibility bridge¹⁴ temporarily installed in the Drapers Field Study Centre 3.5 km from the lake. Volume susceptibility (κ) versus depth graph plots were produced

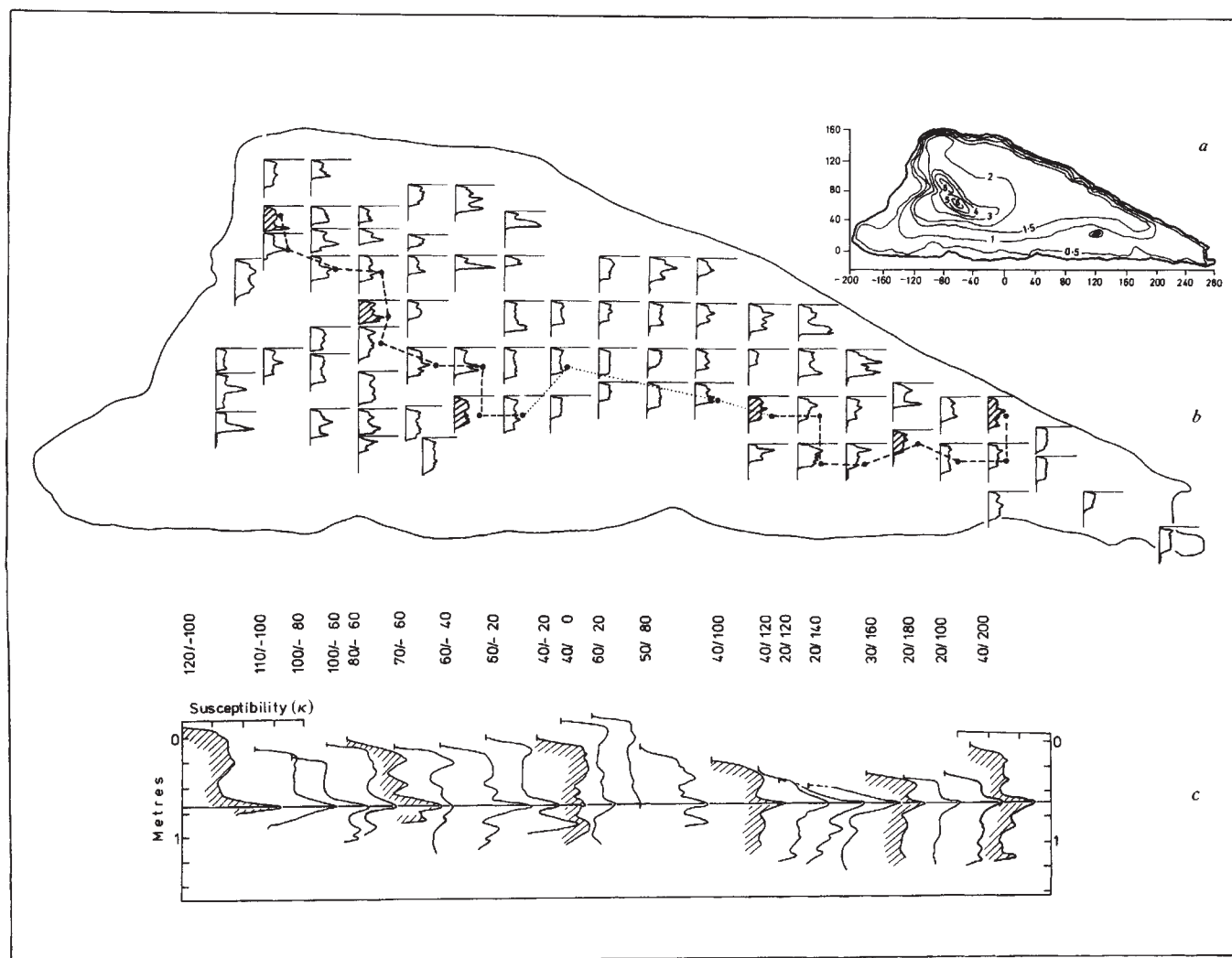


Fig. 1 Llyn Goddionduon: *a*, Bathymetric map and coring grid coordinates; *b*, whole-core susceptibility traces from the main coring grid. The 'tie-line' joins the cores from which the linked and correlated traces shown below in *c* were obtained. The cores for which single sample χ , SIRM and χ /SIRM ratios are plotted in Fig. 2 are identified by shaded traces.

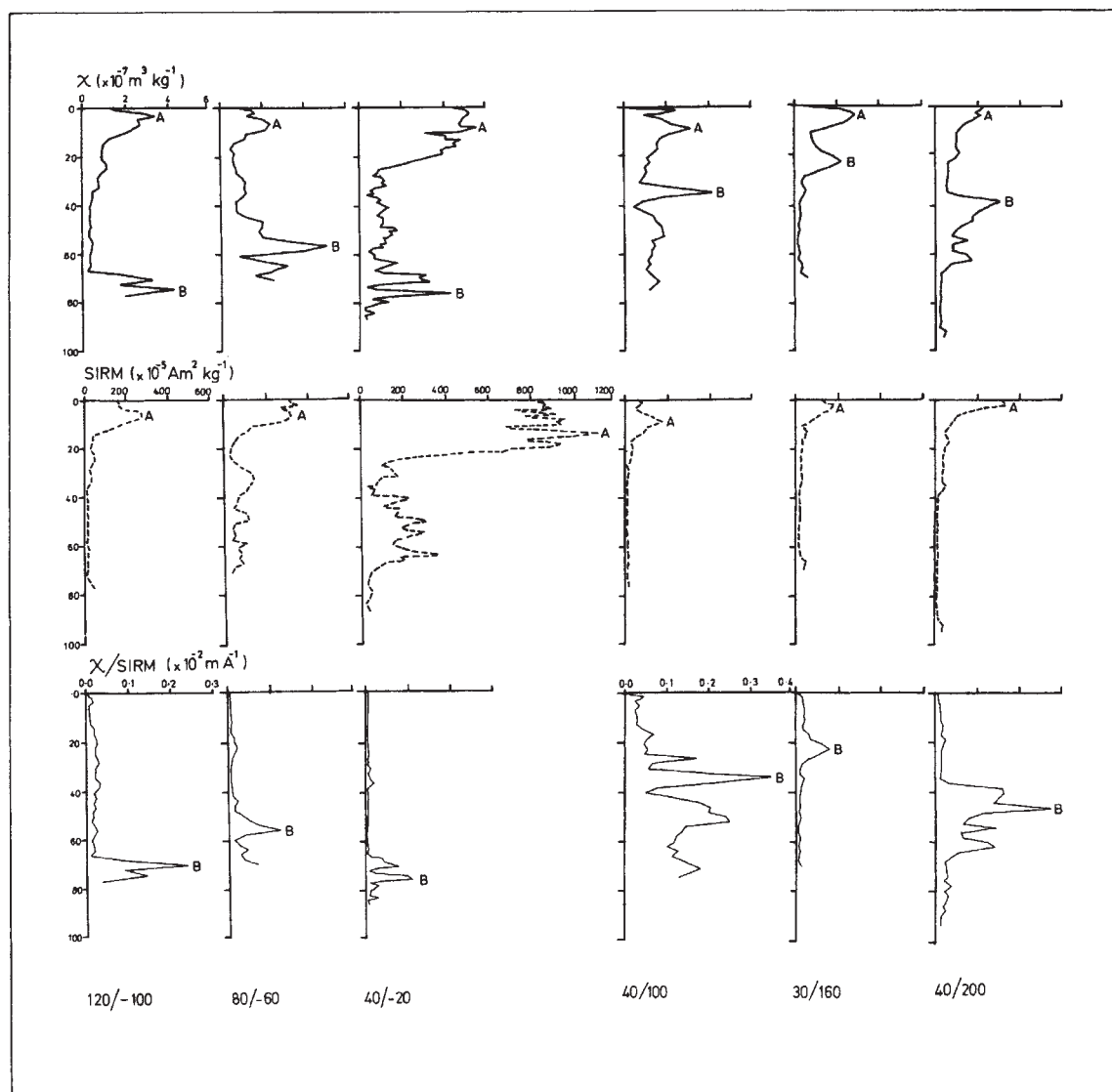


Fig. 2 Llyn Goddionduon. Single sample plots of χ , SIRM and χ/SIRM for six widely separated cores from the main grid. In all cases, the upper peak in χ and SIRM which gives low χ/SIRM values corresponds with the influx of magnetic minerals subsequent to a forest fire in 1951. The lower peak is clearly distinguishable from this in all cores by virtue of its high χ/SIRM ratios.

automatically immediately after sensing and it was thus possible to plan day-to-day field sampling priorities in the light of the results obtained within a few hours of taking each set of cores.

Subsequent studies based on extruded sediment have included measurements of specific susceptibility (χ), the intensity and direction of natural remanence (NRM), anhysteretic remanence (ARM), saturation isothermal remanence (SIRM) and coercivity of SIRM (B_{CR}). These magnetic measurements provide a wide range of rapidly calculated parameters and ratios which may be used to test core-correlation schemes. In addition chemical, particle size, loss-on ignition, pollen-analytical and radiometric studies are in progress.

Figure 1b shows whole core susceptibility traces for a selection of cores from the main grid of cores for the whole lake. In view of the paucity of Flandrian (Post-glacial) sediments in the tiny trench zone (<1 m in all 27 cores from only 5% of the lake), the complex deposition pattern of these cores has little effect on 'whole lake' investigations and the trench results can be satisfactorily assessed on the main 20-m grid. A 'tie-line' has been drawn down the length of the lake. The tie-line cores have been provisionally correlated using the scheme of whole core susceptibility peak matching indicated in Fig. 1c and substantiated in more detail by the plots of specific magnetic parameters from

dried contiguous 1- or 2-cm subsamples presented in Fig. 2. The central zone in Fig. 1b reflects an area of more extended deposition within which the main susceptibility peak used in correlation lies below the depth of sediment penetrated by 1.3-m cores.

Figure 2 shows down core χ , SIRM and χ/SIRM ratios for dried sub-samples in, six widely separated cores located on Fig. 1. Peak χ/SIRM ratios associated with the main tie-line linking the sites, and low χ/SIRM ratios characteristic of the near surface peak (masked by the core tops in whole-core measurements) are common to all the cores and support the proposed correlation scheme. The uppermost sediment with low χ/SIRM ratio is attributable to the forest fire of 1951 (ref. 11) which occurred on the north-east edge of the drainage basin.

Studies in progress are designed to confirm the full pattern of correlation, to date each stage in accumulation, to characterise sedimentologically, chemically and magnetically the variations in deposition rate portrayed, and to relate them to the history of land use at the site decipherable in the pollen record. These studies will provide a basis for empirically based calculations of changing net sediment input versus age for almost all the lake basin for the time interval spanned by the majority of the cores.

Magnetic measurements now open up the prospect of core scanning and correlation with a speed, and on a scale, orders of

magnitude greater than conventional techniques allow. Well founded empirical calculations of total sediment input versus time thus becomes feasible even for lakes as complex as the present example. In the case of Llyn Goddionduon the anticipated pattern of sediment focusing into what is at present the deepest part of the lake has not been characteristic of the Flandrian period. Nevertheless, correlatable patterns emerge from the magnetic studies even within and between zones of relatively shallow marginal deposition.

We thank the NERC, the Leverhulme Trust and the Universities of Liverpool and Edinburgh for financial support, also colleagues and students from both Universities for field assistance, and we are grateful for the facilities in the Draper's Field Centre, Rhyd-y-Creiau.

J. BLOEMENDAL
F. OLDFIELD

Department of Geography,
University of Liverpool,
Liverpool, UK

R. THOMPSON

Department of Geophysics,
University of Edinburgh,
Edinburgh, UK

Received 6 March; accepted 24 May 1979.

1. Battarbee, R. W. *Phil. Trans. R. Soc. B* **281**, 303 (1977).
2. Pennington, W., Cambray, R. S., Eakins, J. D. & Harkness, D. D. *Freshwat. Biol.* **6**, 317 (1976).
3. Mackereth, F. J. H. *Phil. Trans. R. Soc. B* **250**, 165 (1966).
4. Oldfield, F. *Prog. Phys. Geogr.* **1**, 460 (1977).
5. Davis, M. B. *Envir. Conserv.* **3**, 130 (1976).
6. Lehman, J. T. *Quat. Res.* **5**, 541 (1975).
7. Thompson, R., Battarbee, R. W., O'Sullivan, P. E. & Oldfield, F. *Limnol. Oceanogr.* **20**, 687 (1975).
8. Oldfield, F., Dearing, J., Thompson, R. & Garrett-Jones, S. E. *Pol. Arch. Hydrobiol.* **25**, 321 (1978).
9. Oldfield, F., Rummary, T. A., Thompson, R. & Walling, D. E. *Water Reservoir Res.* (in the press).
10. Longworth, G. *et al. J. Soil Sci.* (in the press).
11. Rummary, T. A., Bloemendal, J., Dearing, J., Oldfield, F. & Thompson, R. *Ann. Geophys.* (in the press).
12. Liddle, M. J. *A Survey of Twelve Lakes in the Gwydyr Forest Region of Snowdonia*. (University College of North Wales, 1974).
13. Mackereth, F. J. H. *Limnol. Oceanogr.* **14**, 145 (1969).
14. Molyneux, L. & Thompson, R. *Geophys. J. R. astr. Soc.* **32**, 429 (1973).

²¹⁰Pb dating of annually laminated lake sediments from Finland

USE of ²¹⁰Pb dating is increasing rapidly and applications include studies of accelerated eutrophication in major lakes¹, salt-marsh accretion², the recent history of heavy metal pollution³ and accelerating soil erosion resulting from subsistence agriculture⁴. As dating models have increased in variety and complexity, it is important to compare models against precise and unambiguous independently derived time scales. In each area of application of ²¹⁰Pb dating, the inferences drawn from the calculated age-depth curves and the estimates of changing flux rates are often highly dependent on the ²¹⁰Pb dating model used. In this report ²¹⁰Pb-derived estimates of lake sediment age and dry-mass sedimentation rates are compared with ages and rates calculated directly by counting annual laminations. The results support a model of ²¹⁰Pb dating which assumes a constant net rate of supply (c.p.s.) of unsupported ²¹⁰Pb to the sediment despite fluctuations in dry mass sedimentation rates. Our findings underline the need for empirical evaluation of alternative ²¹⁰Pb dating models in the widest possible range of contexts. They also cast doubt on some published studies in which strongly 'kinked' profiles of unsupported ²¹⁰Pb concentration have been interpreted within the framework of conventional constant initial concentration (c.i.c.) assumptions.

Most workers using ²¹⁰Pb dating have assumed the initial unsupported ²¹⁰Pb concentration to be constant throughout a

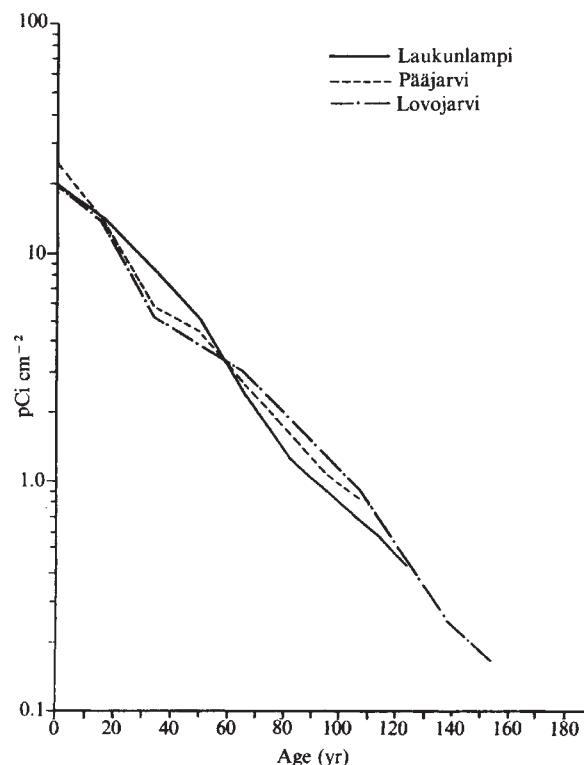


Fig. 1 Cumulative residual unsupported ²¹⁰Pb versus age in three annually laminated sediment profiles from lakes in east Finland. The approximation to a negative exponential function in each case confirms the validity of this as the appropriate dating parameter.

dated profile, or at least in large sections of the profile. This assumption can be satisfied in theory by a direct association between sedimentation and unsupported ²¹⁰Pb supply⁵, or by a combination of constant unsupported ²¹⁰Pb flux and constant dry-mass sedimentation for all of⁶, or segments⁷ of, each profile. Authors using c.i.c. models have dealt with sharply inflected or non-monotonic ²¹⁰Pb-versus-depth profiles, or individual 'anomalous' values, by invoking alternative explanations such as physical mixing⁸, bioturbation⁹, diffusion¹⁰ and co-precipitation of ²¹⁰Pb with manganese¹¹. Where authors have referred to the alternative basic assumption of a constant net rate of supply (c.r.s.) of unsupported ²¹⁰Pb to the sediment independent of fluctuations in bulk sedimentation rate this alternative has usually been dismissed in favour of some variant of the c.i.c. model^{12,13}, or else it has been used but not empirically tested in a rigorous manner^{2,14,15}. Recent applications of c.r.s. model calculations¹⁶ to lake sediment sequences in Northern Ireland, the Highlands of Papua New Guinea¹⁷ and Lake Michigan¹⁸ provide strong, varied, but essentially circumstantial evidence for the validity of c.r.s.-derived age-depth relationships and dry-mass sedimentation rates. In each of these case studies, as in the example by Robbins¹⁵, the choice between alternative models has a major impact on the results and their interpretation.

The sites from which the present evidence has been obtained provide a very precise and unambiguous basis for testing the alternative dating models. Not only are their recent sediments in the form of distinctive annual laminations, confirmed both biostratigraphically^{19,20} and by ^{239,240}Pu and ¹³⁷Cs analyses (T. Jaakola, K. T. P. H. and S. Tiainen, unpublished), but they have also experienced disturbance in each catchment which has given rise to fluctuations in sedimentation rate strong enough to 'kink' the unsupported ²¹⁰Pb concentration-versus-depth curve.

In the c.r.s. model, the age t of sediments of depth x is calculated from¹⁶

$$A(x) = A(0) e^{-kt} \quad (1)$$

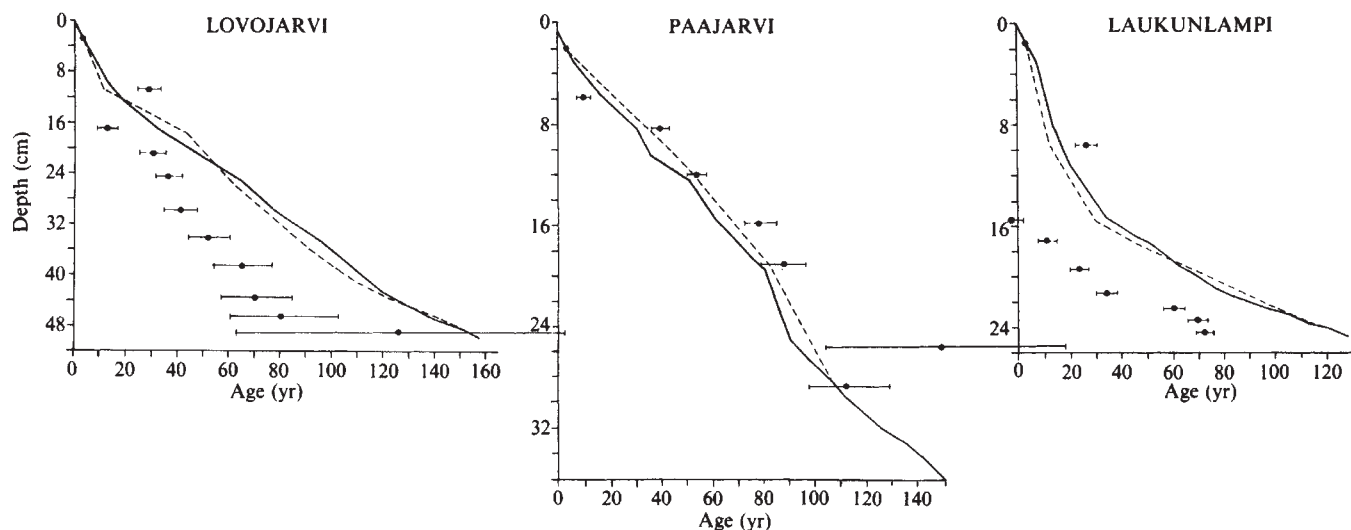


Fig. 2 Age-depth curves based on annual laminae- and on c.r.s.-model ^{210}Pb dating. —, Laminae based curve; ---, c.r.s. model curve; ●, points on c.i.c. model.

where $A(x)$ is the residual unsupported ^{210}Pb beneath sediments of depth x , and k is the ^{210}Pb radioactive decay constant. The values of the dating parameter $A(x)$ for the Finnish lakes were calculated by numerical integration of the unsupported ^{210}Pb concentration with regard to the cumulative dry mass. The residual unsupported ^{210}Pb beneath the lowest datum point was estimated by assuming that the mean ^{210}Pb flux rate during the period covered by the observations was similar to the mean ^{210}Pb flux rate before this period. In Fig. 1, $A(x)$ is plotted against time for all three profiles, confirming the close approximation of this parameter to the theoretical relation expressed by equation (1). There is a close correspondence in the cumulative residual unsupported ^{210}Pb between sites, and hence in the

^{210}Pb flux rates, despite their physical separation, varied morphometry and contrasted sedimentation rates. The location and morphometric characteristics of each lake and drainage basin are summarised in Table 1, which also records mean ^{210}Pb flux rates, and total residual unsupported ^{210}Pb contents of each core.

Figure 2 compares for each site, the age-depth relationship derived from counting annual laminae, with the c.r.s. based ^{210}Pb age-depth profiles. In accordance with the assumption used in estimating the initial value of the dating parameter, the curves are synchronised at the lowest data point. The age-depth points derived using the c.i.c. model are also shown, although no attempt has been made to translate these into an age-depth

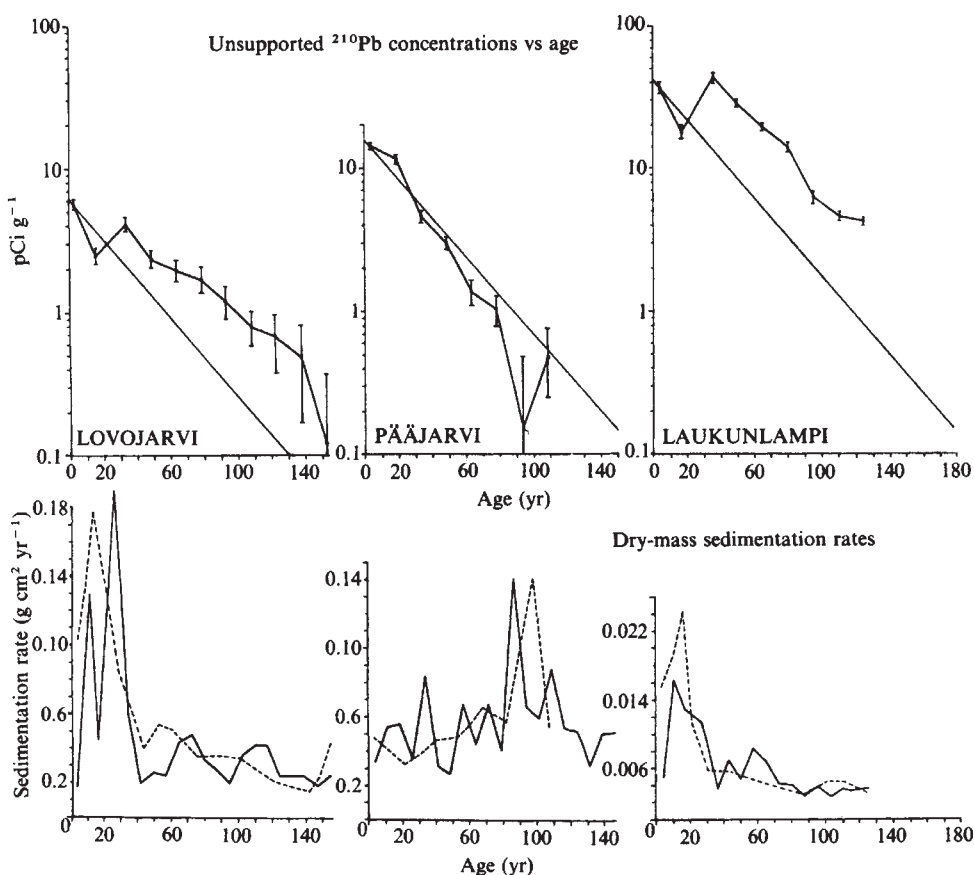


Fig. 3 Unsupported ^{210}Pb concentrations plotted against age (upper) and dry mass sedimentation rates based on annual laminae thickness and on c.r.s. model ^{210}Pb calculations (lower). In each sediment profile, the concentration-versus-age curve deviates significantly from the negative exponential function representing the expected radioactive decay of ^{210}Pb and shown by the straight line. Each major inflection in the concentration-versus-age profile corresponds with peak dry-mass sedimentation rates determined from the annual laminae (—). Sedimentation rates based on c.r.s. model ^{210}Pb are shown as dashed line (---).

Table 1 Morphometric data, mean unsupported ^{210}Pb flux and total residual unsupported ^{210}Pb content

Lake	Area (A, hectare)	Drainage area (D)	D/A	Maximum depth (m)	Volume (m^3)	No. of direct inflows	Mean unsupported ^{210}Pb flux ($\text{pCi cm}^{-2} \text{ yr}^{-1}$)	Total residual unsupported ^{210}Pb content (pCi cm^{-2})
Lovojärvi 61°85' N, 25°92' E	4.8	570 hectare	119.1	17.5	36.9×10^4	6	0.65	20.09
Pääjärvi 61°04' N, 25°05' E	1,342.4	244 km^2	18.18	87	205.7×10^6	25	0.66	24.45
Laukunlampi 62°40' N, 29°10' E	8.4	21 hectare	2.5	26	—	0 (closed basin)	0.64	20.47

relation in view of their generally poor match with true age irrespective of which adaptation of the c.i.c. model is used.

Figure 3 shows changing unsupported ^{210}Pb concentrations plotted against time for each site compared with the dry-mass sedimentation rates derived from both the laminae counts and the c.r.s. calculations. In each case, depressed concentrations correspond with independently verified peak sedimentation rates, as required by the c.r.s. model.

At no point in any profile do the laminae- and c.r.s.-derived ages differ by more than 10 yr. At major points of inflection in each concentration-versus-depth profile, whether the inflections date from 5 to 10 yr or 100 yr before the year of sampling (1977–78), the c.r.s. calculations give an excellent approximation to true age. The close agreement is confirmed despite the need to interpolate between data points, an argument previously used to cast doubt on the validity of c.r.s.-based dating¹³. For dry-mass sedimentation rates, the agreement between laminae-based and c.r.s.-based estimates is less close partly because of the short time scale over which sedimentation rates appear to have varied in comparison with the ^{210}Pb sampling interval, and partly because the c.r.s. calculations of sedimentation rates are intrinsically more sensitive to the effect of experimental error than are the age-depth calculations. Nevertheless, the general characteristics of the changes in sedimentation rate are reasonably well portrayed by the c.r.s. calculations, within the limits of accuracy imposed by the sampling interval. Further studies incorporating mixing models¹⁸, procedures for estimating patterns and rates of sediment focusing from ^{210}Pb measurements, comparison between ^{210}Pb derived ages and precisely dated influx events in non-laminated sediments, and systematic error analysis²¹ are in progress. Treatment of over 40 ^{210}Pb profiles from 18 contrasted lacustrine and near-shore marine environments confirms that when independent tests of internal consistency and external validation can be applied to cases where dry mass sedimentation rates have varied through time, the c.r.s. model provides a better approximation to true age than do any of the alternatives. In no case encountered so far are any of the alternative dating models more accurate when used as a basis for estimating rates of sedimentation or mineral flux.

We thank J. D. Eakins (AERE, Harwell) for carrying out the analyses, and NERC, the Leverhulme Foundation and the Finnish Academy of Sciences for financial support.

P. G. APPLEBY

Department of Applied Mathematics
and Theoretical Physics,

F. OLDFIELD

Department of Geography,
University of Liverpool, UK

R. THOMPSON

Department of Geophysics,
University of Edinburgh, UK

P. HUTTUNEN

Karelian Institute, Section of Ecology,
University of Joensuu, Finland

K. TOLONEN

Department of Botany,
University of Helsinki, Finland

Received 6 March; accepted 30 May 1979.

1. Battarbee, R. W. *Phil. Trans. R. Soc. B* **281**, 303 (1978).
2. McCaffrey, R. J. & Thompson, J. in *Trace Elements in Natural Waters* (ed. Turekian, K. K.) (1974).
3. Goldberg, E. D. *et al. Geochim. cosmochim. Acta* **42**, 1413 (1978).
4. Oldfield, F., Appleby, P. G. & Thompson, R. J. *Ecol. (in the press)*.
5. Pennington, W., Cambray, R. S., Eakins, J. D. & Harkness, D. D. *Freshwat. Biol.* **6**, 317 (1976).
6. Robbins, J. A. B., Edgington, D. N. *Geochim. cosmochim. Acta* **39**, 285 (1975).
7. Brugam, R. B. *Quat. Res.* **9**, 349 (1978).
8. Petit, D. *Earth planet. Sci. Lett.* **23**, 199 (1974).
9. Robbins, J. A., Krezowski, J. & Mozley, S. *Earth planet. Sci. Lett.* **36**, 325 (1977).
10. Koide, M., Bruland, K. W. & Goldberg, E. D. *Geochim. cosmochim. Acta* **37**, 1171 (1973).
11. Megumi, K. *Nature* **274**, 885 (1978).
12. Goldberg, E. D. in *Radioactive Dating* (IAEA, Vienna, 1963).
13. Crozaz, G., Picciotto, E. & De Breuck, W. J. *geophys. Res.* **69**, 2594 (1964).
14. Shackett, W. M. *Rhode Island occas. Publ.* **3**, 29 (1965).
15. Robbins, J. A. *Biogeochemistry of Lead in the Environment* (ed. Nriagu, J. O.) (Elsevier, Amsterdam, 1978).
16. Appleby, P. G. & Oldfield, F. *Catena* **5**, (1978).
17. Oldfield, F., Appleby, P. G. & Battarbee, R. W. *Nature* **271**, 339 (1978).
18. Appleby, P. G. & Oldfield, F. *Envir. Sci. Technol.* **13**, 478 (1979).
19. Simola, H. *Ann. Bot. Fennici* **14**, 143 (1977).
20. Simola, H. *Oikos* (in the press).
21. Battarbee, R. W., Appleby, P. G. & Oldfield, F. *Archs Hydrobiol.* (in the press).

Cinnamic acid inhibition of detritus feeding

AN important advance in ecology has been the general acceptance of Fraenkel's postulate that certain chemicals in plants deter herbivores¹⁻³. Such chemicals are usually termed secondary plant substances because they are not involved in primary metabolic pathways. Very little of the annual production of biomass by higher plants is consumed by herbivores or phytopathogens⁴. Instead, most of the biomass becomes litter and eventually decays through the activity of decomposers. The secondary compounds that deter grazers while the plants are alive do not disappear immediately when plants senesce and die. We have therefore investigated whether these anti-herbivore substances continue to inhibit consumption by organisms feeding on litter or detritus. We report here that cinnamic acids, one type of secondary plant substances found in detritus, inhibit feeding by detritivores. This inhibition occurs at concentrations found in nature and may be a major factor controlling the rate of decay of organic matter.

In salt marshes most of the annual production by plants becomes litter. The most prominent saltmarsh plants in the eastern coast of the US are grasses of the genus *Spartina*. Unlike some grasses⁵, tissues of *Spartina alterniflora* contain few or none of the compounds, such as condensed tannins, alkaloids or cyanogenic glucosides, which are the major deterrents to grazing in many terrestrial plants^{2,6}. It seems likely that in *Spartina* the substances deterring herbivores are the soluble and insoluble esters of glucosides of hydroxyaromatic acids⁷⁻⁹, also found in

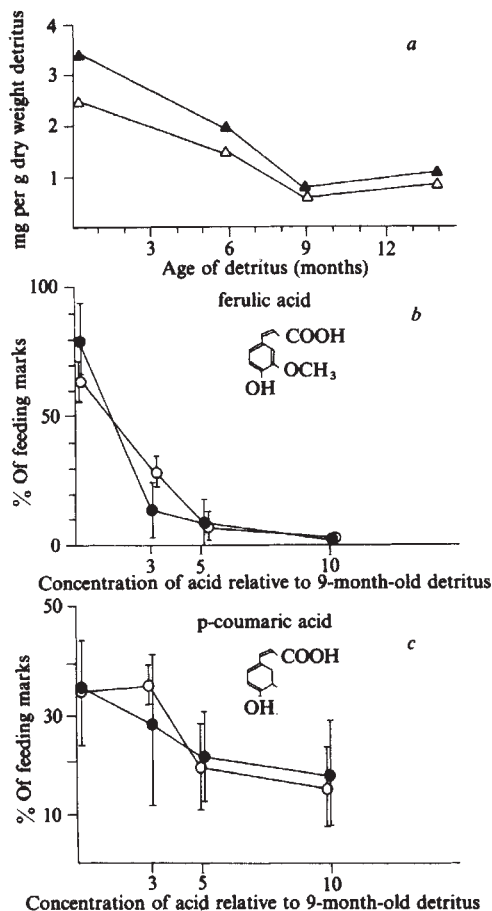


Fig. 1 a, Concentration of ferulic acid (▲) and coumaric acid (△) in litter of *Spartina alterniflora* of increasing age. b, c, Feeding response of *Orchestia grillus* (●) and *Melampus bidentatus* (○) to increased ferulic acid (b) and p-coumaric acid (c) in their food.

Sorghum bicolor^{10,11}. These compounds are readily hydrolysed by hydrolases in the plant to form the actual deterrents, the corresponding free acids^{10,11}. Our work focused on ferulic acid and p-coumaric acid (Fig. 1), cinnamic acids that are abundant in the cell walls of *S. alterniflora*.

We extracted free cinnamic acids from tissues of dead *S. alterniflora* in 80% methanol at 85 °C for 10 min. The extract was cooled for 20 min and filtered through a Millipore filter; the procedure was then repeated. The unhydrolysed extract of detritus of four different ages was analysed by TLC and spectrophotometry for ferulic acid (320 nm) (Fig. 1a). Both acids decreased in concentration with the age of the detritus so that after 9 months of natural ageing in the field only one-third of the initial concentration of cinnamic acids remained in the detritus.

We tested the effect of ferulic and p-coumaric acids on feeding by two saltmarsh detritus feeders, the amphipod *Orchestia grillus* and the snail *Melampus bidentatus*. We suspended ground, dead grass in agar in a Petri dish with four compartments. This agar suspension method of testing palatability was developed by Gieselmann¹². The amphipods and snails fed very well on the suspension and both species fed on the agar only when detritus was present.

We tested the effect of ferulic and p-coumaric acids by adding them to the agar suspensions of 9-month-old detritus of *S. alterniflora*, plating the suspensions, allowing the test animals to feed. The detritus was aged naturally in the field. We added 0, 3, 5 and 10 times the amounts of acids found in 9-month-old detritus, and each of these four doses was included in every Petri dish. To each of these dishes we introduced four amphipods or snails. The dishes were incubated in the dark for 3 h. Feeding

was assessed by the number of feeding marks left on the surface of the agar suspension. Marks were easily counted and clearly reflected the difference in mouth parts and feeding mechanisms of the two species (Fig. 2).

An increase in the concentration of both ferulic and p-coumaric acids decreased feeding activity by both *O. grillus* and *M. bidentatus* (Fig. 1b, c). Ferulic acid was more effective than p-coumaric acid. When feeding on recently dead plant tissues, the detritus feeders are exposed to concentrations of these acids up to three times that found in 9-month-old detritus (Fig. 1a). We can expect, therefore, that detritus of 9 months or older could be significantly more palatable than new detritus due to the lower content of ferulic acid. Concentrations of more than five times the amount of p-coumaric acid in 9-month-old dead grass reduced feeding. Because we did not find such high concentrations in litter or detritus of *S. alterniflora*, p-coumaric acid may not be as important as ferulic acid in deterring detritus feeders.

The mechanism that makes ferulic acid distasteful is unclear. Detritus feeders must taste the acidity as well as other flavours of the compound. In preliminary tests with ferulic acid buffered at neutral pH amphipods, but not snails, were inhibited, suggesting that some detritus feeders are deterred by acidity, whereas others are inhibited by sour or bitter tastes.

Other investigators have looked at microbial density and carbon and nitrogen contents as providing different ways of measuring the food quality of aquatic detritus^{13,14}. Our evidence suggests that secondary plant substances such as ferulic acid are important in determining whether organic matter is consumed by detritus feeders. Further, because there are marked changes in the amounts of these compounds in litter as it ages and becomes detritus, the time at which food may become palatable

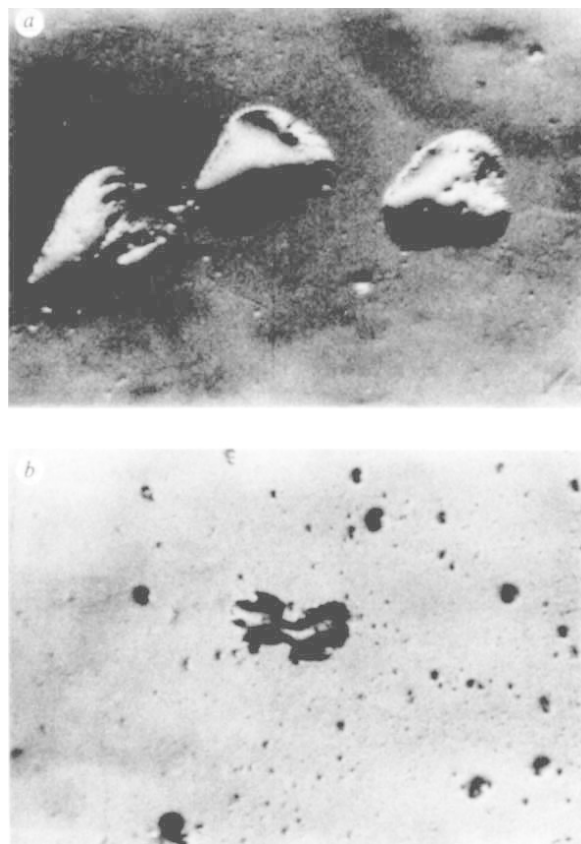


Fig. 2 Feeding marks of *Melampus bidentatus* (a) and *Orchestia grillus* (b) on the surface of an agar-detritus suspension. The marks are about 0.4 mm for *M. bidentatus* and 0.3 mm for *O. grillus* at their longest dimension. The radular marks left by the snail (*M. bidentatus*) are single whereas the amphipod (*O. grillus*) leaves double marks as a result of biting with paired mandibles.

to decomposers may also be determined by secondary plant substances.

The rate of decay of organic matter is accelerated by the comminuting of detritus by detritivores. The smaller particles provide new surfaces for microbial attack. Further, passage of detritus through the digestive system of detritivores also alters the nature of the particles, facilitating microbial reinvasion and also accelerating rates of decay. Secondary plant substances present in detritus, by inhibiting feeding by detritivores, thus delay the rate at which organic matter decays. It is also possible that microbial activity itself may be inhibited by such substances. This needs study, but it is clear that cinnamic acids, inherited from live plants, and perhaps other compounds, have a critical role in determining rates of decomposition of detritus. Thus, secondary plant substances may be a major control mechanism in ecosystems such as marshes where the detritus food web dominates the flow of carbon.

This work was supported by NSF grants DEB-7683877 and SPI 76 83465A01. We thank Dorothea and the late Arnold Gifford and Salt Pond Sanctuaries for use of their land for this work. The photographs in Fig. 2 are by Ana Sylvester Serianni.

I. VALIELA
L. KOUMJIAN
T. SWAIN

Boston University Marine Program,
Marine Biological Laboratory,
Woods Hole, Massachusetts 02543
and

Department of Biology,
Boston University,
2 Cumming Street,
Boston, Massachusetts 02215

JOHN M. TEAL

Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543

JOHN E. HOBBIE

Ecosystems Center,
Marine Biological Laboratory,
Woods Hole, Massachusetts 02543

Received 19 March; accepted 15 May 1979.

1. Fraenkel, G. F. *Science* **129**, 1466–1470 (1959).
2. Whittaker, R. H. & Feeny, P. P. *Science* **171**, 757–770 (1971).
3. Swain, T. A. *Rev. Pl. Physiol.* **28**, 479–501 (1977).
4. Feeny, P. in *Coevolution of Animals and Plants* (eds Gilbert, L. E. & Raven, P. H.) Ch. 1 (University of Texas Press, Austin, 1976).
5. Hegnauer, R. *Chemotaxonomie der Pflanzen* Vol. 2 (Birkhauser, Basle, 1963).
6. Levin, D. & York, B. M. *Biochem. Systematics Ecol.* **6**, (1979).
7. Harborne, J. B. & Williams, C. A. *Biochem. Systematics Ecol.* **4**, 267–280 (1976).
8. Swain, T. in *Herbivores: Their Interactions with Secondary Plant Metabolites* (eds Rosenthal, G. A. & Jansen, D. W.) (Academic, New York, in the press).
9. Hartley, R. D. & Jones, E. C. *Phytochemistry* **15**, 1157–1160 (1976).
10. Woodhead, S. & Bernays, E. A. *Entomologia exp. appl.* **24**, 123–144 (1978).
11. Woodhead, S. & Cooper-Driver, G. A. *Biochem. Systematics Ecol.* **7**, (1979).
12. Gieselmann, J. A. 41st A. Ming Am. Soc. Limnol. Oceanogr. Astr. (1978).
13. Tenore, K. R. *Limnol. Oceanogr.* **22**, 936–941 (1977).
14. Harrison, P. G. & Mann, K. H. *Limnol. Oceanogr.* **20**, 924–934 (1975).

Isolation and cultivation of diatom symbionts from larger Foraminifera (Protozoa)

THERE has been increasing interest in larger Foraminifera that have symbiotic algae which apparently adapt them to thrive in certain shallow, nutrient-poor tropical seas where they are responsible for a significant or even dominant fraction of the carbon fixed in the benthos^{1,2}. Studies of fine structure of several species of these giant protozoa have suggested that their algal symbionts are diatoms^{6–12} without frustules (shells). This suggestion is itself remarkable because almost all marine

invertebrate–brown algal associations (for example, those of corals and tridacnid clams) involve dinoflagellates. Unfortunately, knowledge of the architecture of the frustule is needed for the identification of a diatom. However, because the formation of the outer envelopes of the dinoflagellate and chlorophyte endosymbionts of other large species of Foraminifera is repressed within the hosts^{6,7,13} but not in culture^{5,14}, we hoped that if the diatom endosymbionts could be isolated and cultivated they would form characteristic frustules. We report here the successful isolation and cultivation of diatom symbionts from foraminifera.

With the aid of SCUBA, *Heterostegina depressa*, *Amphistegina lessonii*, *A. lobifera* and *A. papillosa* were collected at 20–35 m in the Gulf of Elat (Red Sea) near Geziret Faraoûn anchorage. Individual animals were placed immediately in sterile dishes and under a dissecting microscope they were meticulously and aseptically brushed, scraped and washed through successive baths of antibiotic-containing sterile seawater to remove all external algae^{5,14}. After processing, each animal was examined with a water immersion lens of a compound microscope to assure removal of all external algae. Individual animals were transferred aseptically to test tubes of sterile seawater and crushed carefully with sterile glass rods to release the symbionts. The algae were then inoculated into isolation media¹⁵.

Growth of algal colonies was observable under a dissecting microscope after 4–8 d. The growth of large numbers of colonies of the same types¹⁵ from each host is strong evidence that the algae isolated were not contaminants. If the isolates were undigested food organisms, then their hosts were filled with them. Growth of transfers was poor and restricted to media containing vitamins B₁₂, thiamine, biotin or mixtures of vitamins (media 15, 17 and 18)¹⁵. The isolates are now being maintained in a natural seawater medium (~40‰) enriched with: 10 mg % (w/v) yeast extract, 1 ml % (v/v) soil extract, 0.1 µM vitamin B₁₂, 1 µM biotin, and 1 µM thiamine.

All isolates were small pennate diatoms (3–14 µm), and two were previously undescribed. Some Foraminifera contained two

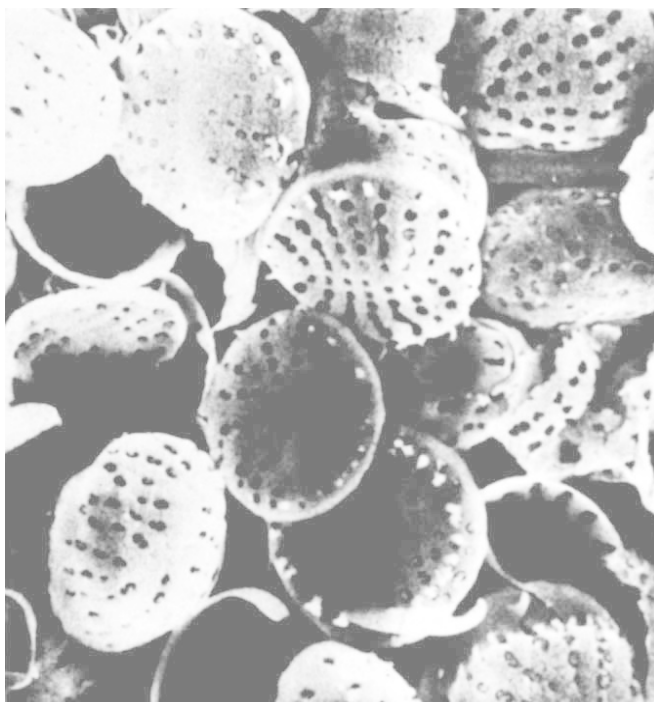


Fig. 1 A new species of zooxanthellae, *Fragilaria* sp., isolated from specimens of the large Foraminifera, *Amphistegina lessonii*, *A. lobifera* and *A. papillosa* collected in the Gulf of Elat (×8,100).

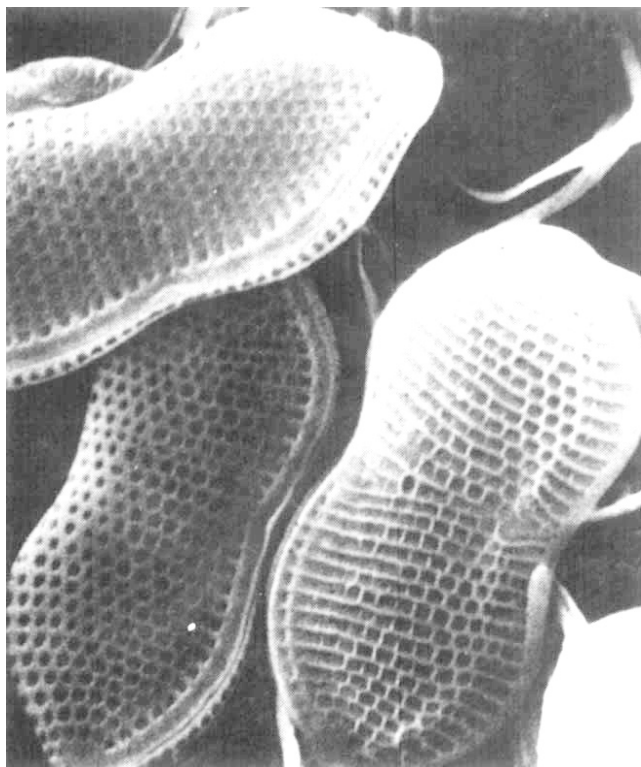


Fig. 2 *Nitzschia panduriformis* v. *continua* Grun. isolated from all specimens of the larger foraminiferan *Heterostegina depressa* and a few specimens of *Amphistegina lobifera* collected in the Gulf of Elat ($\times 5,575$).

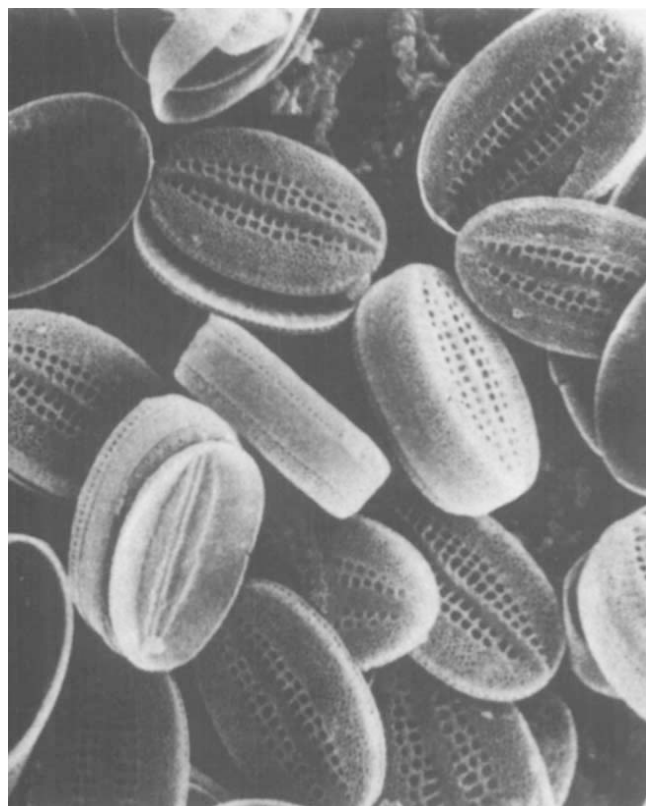


Fig. 3 A new naviculoid diatom isolated from some specimens of *Amphistegina papillosa* collected in the Gulf of Elat ($\times 2,400$).

or rarely three different diatom species. A very small new species of *Fragilaria* (Fig. 1) was found in specimens of all three species of *Amphistegina*. *Nitzschia panduriformis* v. *continua* was found in *Heterostegina depressa* and a few *A. lobifera* (Fig. 2). Several *A. papillosa* contained an unusual new naviculoid diatom (Fig. 3). *Amphora tenerrima*, perhaps an episymbiont, was isolated from the exterior surfaces of both *H. depressa* and *Amphisorus hemprichii*. A manuscript containing full descriptions of the new diatoms is in preparation. None of the symbiont-diatom species was abundant in the epiphytic community at the time of collection.

Traces of chlorophyll *b* in extracts of pigments from intact animals suggested that *Amphistegina* spp. and *Amphisorus hemprichii* also contained symbiotic zoochlorellae. We isolated and grew clones of green algae from them.

The isolation of these symbionts paves the way for physiological-ecological studies which will advance our understanding of the carbon budgets and depth distributions of their hosts. When it is possible to obtain aposymbiotic hosts without killing them, we should be able to satisfy Koch's postulates and explore questions of host specificity and host-symbiont interactions. Research along these lines will be particularly exciting because algal endosymbiosis may have been a driving force in the evolution of these giant protozoa and a key to their great success.

This work was supported by grants from the NSF and the Binational Science Foundation. We thank Drs L. Hottinger, S. Leutenegger and Dr A. Svoboda for collecting most of the specimens; Dr Y. Cohen for help with organisation and Ms Judith Garrison, Mr Carmine Mastropaolo and Robert Koestler for technical assistance.

JOHN J. LEE
MARIE E. MCENERY

Department of Biology,
City College of CUNY,
New York, New York 10031

MOISHE SHILO
Department of Microbiological Chemistry

ZEEV REISS
Department of Geology,
Hebrew University,
Jerusalem, Israel

Received 9 April; accepted 24 May 1979.

1. Sournia, A. *Mar. Biol.* **37**, 29-32 (1976).
2. Sournia, A. *Int. Rev. Ges. Hydrobiol.* **62**, 813-819 (1977).
3. Ross, C. *Proc. Second int. Coral Reef Symp.* **1** Great Barrier Reef Committee. Brisbane (October) 327-333 (1974).
4. Lee, J., McEnery, M., Kahn, E. & Shuster, F. *Micropaleontology* **25**, 118-140.
5. Lee, J. in *Biochemistry and Physiology of Protozoa* (eds Hutner, S. & Levandowsky, M.) (Academic, New York, in the press).
6. Leutenegger, S. *Utrecht. Micropaleont. Bull.* **15**, 225-239 (1977).
7. Leutenegger, S. *Cahiers de Micropaleont.* **3**, 5-53 (1977).
8. Leutenegger, S. *Mar. Biol.* **44**, 157-164 (1977).
9. Dietz-Eibrachter, G. *Meteorol. Forsh. Ergebnisse* **6**, 41-47 (1971).
10. Schmaljohann, R. & Röttger, R. *Naturwissenschaften* **63**, 486 (1976).
11. Hansen, H. & Buchardt, B. *Utrecht. Micropaleont. Bull.* **15**, 205-224 (1977).
12. Berthold, W. *Arch. Protistenk.* **120**, 16-62 (1978).
13. Müller-Merz, E. & Lee, J. *J. Protozool.* **23**, 390-396 (1976).
14. Lee, J., Crockett, L., Hagen, J. & Stone, R. *Br. phycol. J.* **9**, 407-422 (1974).
15. Lee, J., McEnery, M., Kennedy, E. & Rubin, H. *J. Phycol.* **2**, 14-49 (1975).

Changes in the circuitry of the kitten visual cortex are gated by postsynaptic activity

MOST neurones in the visual cortex of normal adult cats are binocular and respond very selectively to elongated contours of a particular orientation¹. If a kitten is deprived of vision in one eye during a critical period in early life the afferents from this eye lose the ability to excite cortical cells². This functional deafferentation is, however, reversible if, still within the critical

period, the previously open eye is closed and the other opened³. These changes in afferent connectivity are attributed to competitive interactions between the pathways from the two eyes^{4,5}, which have been shown to occur at the level of the striate cortex⁶. We previously demonstrated⁷ that afferents from the deprived eye remain functional when activity from the open eye is enhanced but fails to drive the common cortical target cell. This had suggested that the competitive suppression of one subset of converging afferents requires, in addition to imbalance in presynaptic activity, that the postsynaptic neurones actually respond to the more active pathways. Such postsynaptic gating was first postulated by Hebb⁸, and since then it has been assumed in most models on adaptive neuronal connections. In the study reported here more direct evidence for the participation of postsynaptic factors both in competitive suppression and in functional recovery of converging afferents was obtained. The results support Hebb's early hypotheses and demonstrate that both changes in connectivity are in fact guided by postsynaptic response properties.

Three kittens were dark reared until the age of 5 weeks and then, with one eye closed, exposed to a normal environment for 9 consecutive days. Control experiments and a number of previous studies have shown that by this time most cortical cells have become dominated by the open eye and display orientation selectivity comparable to adult standards. At this age the initial shift in ocular dominance can still be reversed when the previously open eye is closed and the deprived eye opened³. If this functional reorganisation of afferent connectivity required responses from postsynaptic cells the following prediction can be made: when the spectrum of orientations visible by the open eye is restricted to a narrow range, reversal in ocular dominance should occur only for those neurones whose response properties enable them to react to the range of visible orientations. Cells with other orientation preferences that have no predisposition to respond to the signals from the open eye should remain connected to the eye that was open first and had experienced the normal environment.

We used artificial astigmatism⁹ to produce effective restriction of the range of visible orientations during this second exposure period. A cylindrical lens of -25 dioptres was mounted in a polyurethane mask and fitted in front of the initially deprived eye while the other, previously open, eye was occluded. Photographs taken through the lens of the environment in which the kittens were exposed confirmed that only orientations within $\pm 10^\circ$ of the zero axis of the lens remained visible. Apart from the minimal counter-rotation of the eyes ($\sim 10\%$ of tilt angle¹⁰) this range remains constant with respect to retinal coordinates regardless of head and body movements. The kittens were therefore allowed to freely explore their environment, which contained high-contrast contours of all orientations. In two kittens vision was restricted to horizontal orientations and in one to vertical orientations. After 10 days of 10 h exposure daily (during the night the kittens were returned to the dark room) the kittens were again kept in the dark for a few days until the experiment. For convenience the eye that was open first is referred to here as the 'first eye', and the eye that had seen only the restricted range of orientations is referred to as the 'second eye'.

The preparation and recording techniques were conventional and are described elsewhere¹¹. The orientation selectivity and the ocular dominance of single neurones in area 17 were determined from manual field plots and from additional averaged response histograms.

As expected, after sequential monocular occlusion, the large majority of neurones were monocular (Fig. 1*a*); but the reversal of ocular dominance was incomplete. The second eye had regained dominance only over those cells whose receptive field orientations matched the visible spectrum of orientations (Fig. 1*b*). In these cells the afferents from the first eye had become ineffective; they remained, however, efficient in neurones whose orientation preference did not correspond to the range of orientations seen by the second eye (Fig. 1*c*). This indicates that

circuit changes had occurred only for those neurones whose functional predisposition had enabled them to react to the activity from the second eye.

This conclusion is further supported by the distribution of orientation preferences of cells in different ocular dominance classes. Comparison of Fig. 1*e* and *f* shows that these polar distributions are complementary to each other. As expected, the cells which are now dominated by the second eye (ocular

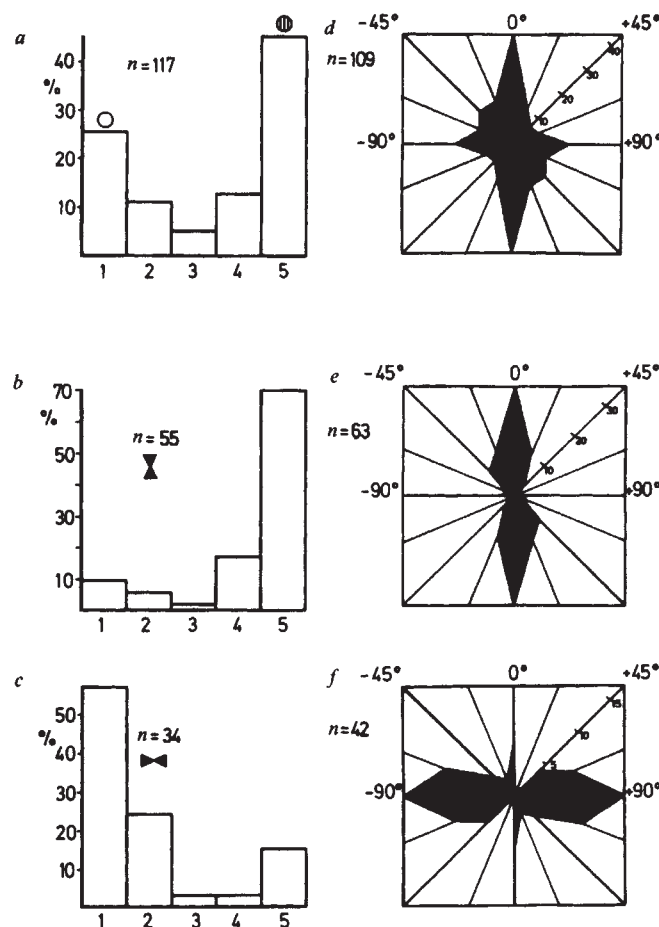


Fig. 1 Ocular dominance (OD) distributions (*a-c*) and orientation preferences (*d-f*) of neurones in the visual cortex of three kittens after reversal of monocular deprivation and selective exposure of the newly open eye to a narrow range of orientations. Number of neurones analysed, 176; those responsive to light, 127 (72%); those visually unresponsive, 49 (28%). Both ocular dominance and orientation selectivity fully determined, 117. Orientation selective or orientation biased, 109 (93%); non-orientated, 8 (7%). Ocular dominance distributions are plotted separately: *a*, for all neurones together irrespective of their orientation preference; *b*, for neurones with orientation preferences corresponding to the orientation ($\pm 22.5^\circ$) experienced by the second eye; and *c*, for neurones with orientation preferences orthogonal ($\pm 22.5^\circ$) to those experienced during the second stage of exposure. OD classes 1 and 5 refer to monocular cells excitable exclusively from the originally open eye (class 1) and the eye that was open last and viewed a narrow range of orientations (class 5). OD class 3 comprises binocular cells that are equally responsive to either eye and classes 2 and 4 contain binocular cells in which the first or second eye, respectively, is dominant. Polar plots of preferred orientations: *d*, independent of ocular dominance; *e*, for neurones driven monocularly or predominantly by the second eye; *f*, for neurones driven monocularly or predominantly by the first eye. The radius of the graph at each intersection gives the number of cells with that particular orientation preference. Angles from 0 to $\pm 90^\circ$ correspond to the difference between the orientation that was seen by the second eye and the neurones' preferred orientations.

dominance classes 4 and 5, Fig. 1e) have orientation preferences that match the narrow range of orientations visible during the second stage of exposure. In addition, cells with these properties are under-represented in the population of cells still dominated by the first eye (ocular dominance classes 1 and 2, Fig. 1f). This complementarity in orientation preference indicates that: (1) cells now dominated by the second eye were recruited from the pool of neurones initially dominated by the other eye; and (2) this recruitment must have been guided by the response properties of the cortical cells recruited.

These results all support the hypothesis that both suppression and enhancement of excitatory transmission in converging pathways are gated by a matching process between the pattern of presynaptic activity and the response properties of the common postsynaptic target cell. Thus, for the present paradigm, Hebb's⁸ classical postulates on adaptive synaptic connections appear to be fully confirmed. Transmission is enhanced selectively in those pathways that are capable of exciting the postsynaptic target and are contingently active with the latter. On the other hand, those afferents which fail to excite the postsynaptic target become suppressed while the target is activated by other converging afferents. Such a mechanism implies then, that activity from the eyes can only select from among pre-existing connections and only within the limits set by the response properties of the postsynaptic targets. Moreover, as functional recovery after reversal was guided by postsynaptic responses, as apparent from the complementary orientation distributions, one has to infer that the initial disconnection has not been completed by the time of reversal. In view of the short period of preceding monocular deprivation and the rapid recovery after reversal it seems possible that, at this early stage of monocular deprivation, synaptic transmission through the deprived pathway is only weakened while the connections themselves persist. This would allow at least sub-threshold activation or even occasional spiking of the cortical target cells in response to stimulation of the deprived eye. Such activation would then act as the postsynaptic signal for initiating the re-establishment of fully functioning synaptic connections.

While this manuscript was in preparation Cynader and Mitchell⁶ reported experiments in which previously inexperienced kittens were exposed with one eye to a restricted range of orientations while the other was allowed to view normally. They found orientation-dependent changes in ocular dominance and therefore concluded that the competitive interactions between afferents from the two eyes occur at the cortical level. Our present data fully support this conclusion and advance the results of Cynader and Mitchell in that they provide the neuronal mechanism through which the response properties of postsynaptic target cells gate the consolidation and the disruption of converging afferents. The generality of the mechanism proposed in this report is emphasised by the fact that it not only predicts correctly the disruption of pathways observed by Cynader and Mitchell but also, as described above, the selective recovery of previously inactivated connections.

J. P. RAUSCHER
W. SINGER

Max-Planck-Institut fuer Psychiatrie,
Kraepelinstrasse 2,
8000 München 40, FRG

Received 15 December 1978; accepted 18 May 1979.

- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **160**, 106–154 (1962).
- Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **26**, 1003–1017 (1963).
- Blakemore, C. & Van Sluyters, R. C. *J. Physiol., Lond.* **237**, 195–216 (1974).
- Guillery, R. W. *J. comp. Neurol.* **144**, 117–127 (1972).
- Sherman, S. M., Guillery, R. W., Kaas, J. H. & Sanderson, K. S. *J. comp. Neurol.* **148**, 1–18 (1976).
- Cynader, M. & Mitchell, D. E. *Nature* **270**, 177–178 (1977).
- Singer, W., Rauschecker, J. P. & Werth, R. *Brain Res.* **134**, 568–572 (1977).
- Hebb, D. O. *The Organization of Behaviour* (Wiley, New York, 1949).
- Freeman, R. D. & Pettigrew, J. *Nature* **246**, 359–360 (1973).
- Grüsser, O. J. & Grüsser-Cornehl, U. *Pflügers Arch. ges. Physiol.* **270**, 227–238 (1960).
- Singer, W. *Expl. Brain Res.* **50**, 25–41 (1977).

Orientation of horizontal cell axon terminals in the streak of the turtle retina

NEURONES which respond preferentially to slits of light of a given orientation have been found at the ganglion cell level of the retina¹, as well as at higher centres of the visual system^{2–4}. Present knowledge about the morphology of such cells and the organisation of the inputs to these cells does not account for the origins of their specialised receptive field properties. We report here that the horizontal cells impaled in the visual streak of the turtle retina (a specialised linear area centralis⁵ analogous to the fovea of higher vertebrates) respond preferentially to slits of light oriented parallel to the streak and thus parallel to the horizontal meridian of the animal's visual field⁵. Anatomical studies demonstrate that these horizontal cells are all elongated and oriented in the direction of the streak. Thus, our findings show that a class of neurones exists in the retina whose morphologies may form the basis for their specialised receptive field properties.

The techniques used for intracellular recording from neurones in the eyecup of the turtle, *Pseudemys scripta elegans*, and the criteria for horizontal cell identification, have been described previously^{6–8}. The light stimulus was a slit, 0.1 mm wide and 3.0 mm long. The centre of the impaled unit's receptive field was determined as the position yielding the largest response to flashed slits of light oriented along two mutually perpendicular axes.

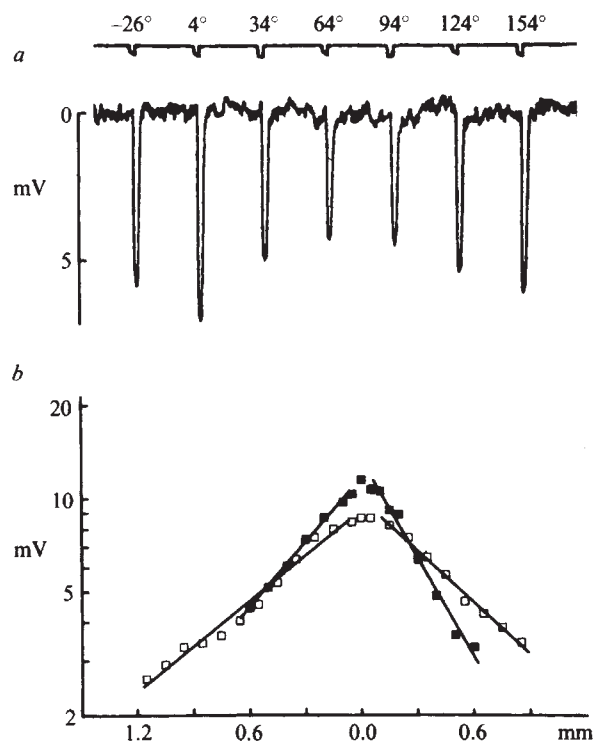


Fig. 1 Receptive field asymmetries in an L_1 -type horizontal cell in the visual streak of turtle retina. *a*, Photoreponses to slits of dim light centred over the receptive field and flashed at 6 s intervals with the orientations given on top of the figure. Angles were measured with respect to the streak orientation (defined as 0°). *b*, Semilogarithmic plot of the peak amplitudes of photoreponses to slits of light as a function of displacement of the slit along directions parallel (□) and orthogonal (■) to the streak. Lines drawn through data points are the best fit to an exponential as determined by a least squares criterion. The slit intensity in B was 3 times that used in A. The maximum response elicited in this unit with bright, 3.2 mm diameter spot illumination was 40 mV.

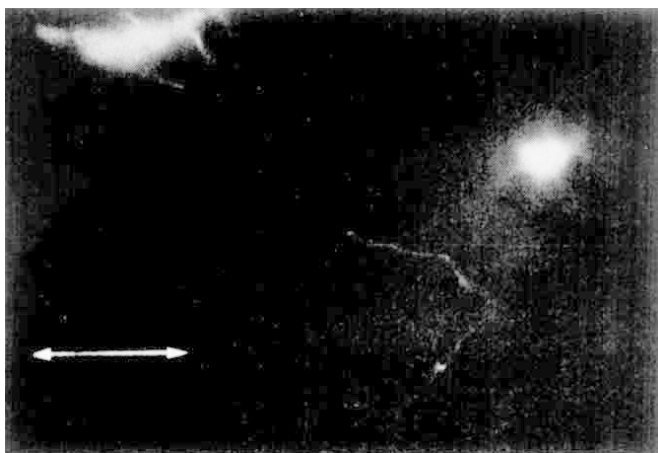


Fig. 2 Lucifer-stained horizontal cell from which intracellular responses of Fig. 1 were recorded. Lucifer has stained the axon terminal (upper left), axon and cell body. The recording site was the axon terminal as indicated by Fast Turquoise marking. Streak direction and scale are indicated by arrow. Scale bar, 30 μm .

The receptive field properties of seventeen luminosity type horizontal cells (L_1 -type) lying within 75 μm of the visual streak of the turtle retina were studied. A slit of dim light intensity was centred over the unit's receptive field and flashed at various orientations. Typical responses are shown in Fig. 1*a*. In this unit, as in all others, the smallest and largest responses were elicited by slits of light oriented approximately perpendicular and parallel to the streak, respectively. The ratio of the smallest to the largest amplitudes recorded (the response eccentricity) was 0.59 for the unit of Fig. 1*a* and the average for all seventeen units was 0.66 ± 0.1 (s.d.).

This non-isotropy in the receptive fields was further investigated in eight units by measuring space constants⁹ in directions parallel and orthogonal to the streak. Figure 1*b* shows a semi-logarithmic plot of peak voltage against slit displacement. The response evoked when the slit was centred over the unit (0.0 in Fig. 1*b*) and parallel to the streak (filled squares) was larger than that evoked by a slit orthogonal to the streak (open squares). Furthermore, the decay of potential with distance from the centre of the receptive field was approximately exponential along each axis and the fall off was greater for slit displacements in the direction perpendicular to the streak (filled squares). The space constants measured on each side of the unit's receptive field centre were different and were averaged to give a mean space constant for each direction. For the unit of Fig. 1*b*, the mean space constants for displacements orthogonal and parallel to the streak were 535 μm and 831 μm , respectively. The ratio between the two space constants provides an index of the asymmetry of the receptive field, and was 0.64 for this unit and 0.73 ± 0.09 on the average.

Lamb⁹ has shown that for dim intensities, the response V , evoked by a slit of width a , centred over a horizontal cell receptive field is related to the space constant λ , by the function $V = E(1 - \exp(-a/2\lambda))$ where E is the response of the cell to a full field stimulus. In this model, λ is related to the sheet and membrane resistances, R_s and R_m , respectively, by $\lambda = (R_m/R_s)^{1/2}$. We have modelled the horizontal cell network in the streak as a two dimensional conductive sheet where the sheet resistance differs in directions parallel and perpendicular to the streak. For this non-isotropic model, the ratio of the voltages calculated for λ values parallel and perpendicular to the streak, $V_{\perp}/V_{\parallel}$, gives an expected response eccentricity which can then be compared to the measured response eccentricity. For the conditions of the experiments described here, $a/2\lambda \ll 1$, so this ratio can be approximated by $V_{\perp}/V_{\parallel} \approx \lambda_{\parallel}/\lambda_{\perp}$. Because the measurements of response eccentricities and space constants were both functions of intensity, data obtained with the same intensity slit were used for a meaningful comparison. The mean difference between predicted and measured response eccentricities was 0.01 ± 0.08 . These findings support the view that the receptive field asymmetries of L_1 horizontal cells as revealed by measurements of space constants and response eccentricities reflect summation within a linear, but anisotropic network where the attenuation of signals is greater in the direction perpendicular to the streak than parallel to it.

Intracellular injection of a mixture of Fast Turquoise¹⁰ (1.2%) and the fluorescent dye, Lucifer¹¹ (3%), was attempted in all seventeen L_1 units and an example is shown in Fig. 2. All six successful stains revealed axon terminals which were elongated and oriented parallel to the streak. In some instances, a faintly stained axon and cell body were also visible (Fig. 2). Luminosity type horizontal cells in the turtle, as in other species, have an axon terminal and a cell body which are connected by a fine axon¹¹⁻¹³ but which are electrically independent^{7,12}. The brighter stains of axon terminals, and the Fast Turquoise marking indicate that the L_1 units are axon terminals.

Anatomical studies of Golgi-Colonnier impregnated horizontal cells in the streak region of the turtle retina provided a morphological correlate for the physiological findings described above. All Golgi-impregnated axon terminals of horizontal cells in the streak region were observed to be elongated and oriented parallel to the streak as shown in Fig. 3. Each axon terminal consisted of a thick, tuberos process 8 μm in diameter and 50 μm in length with numerous small terminals projecting to the photoreceptors. Each axonal arborisation occupied an approximately rectangular space on the retina, with mean dimensions of $20 \times 50 \mu\text{m}$. Also shown in Fig. 3 is a cell body of a horizontal cell with its fine axon.

The actual dimensions of the axonal arborisations (Fig. 3) cannot account for the large receptive fields measured in L_1 units of the turtle streak. There is considerable physiological^{7,9,14-18} and anatomical¹⁹⁻²³ evidence that horizontal cells in the vertebrate retina are extensively electrotonically coupled via gap junctions, thereby forming a conductive sheet across the retina. The elongated shape of the axon terminals in the turtle streak, and their common orientation parallel to the streak suggests that

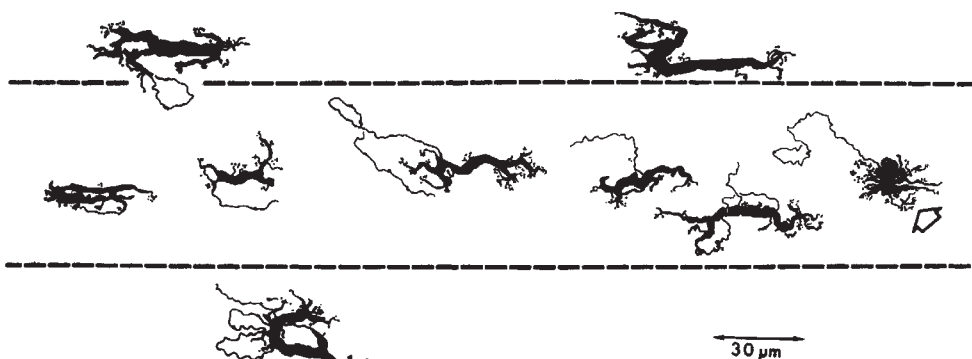


Fig. 3 Drawings of Golgi-Colonnier impregnated horizontal cell axon terminals in the visual streak region (area between dashed lines). The axon terminals are connected via the thin curly axon to horizontal cell bodies of the type illustrated (arrow). The axon terminals have arborisations elongated in the direction of the streak.

the mean distance between gap junctions linking neighbouring L_1 units should be greater in the direction parallel to the streak than perpendicular to it. Assuming that cytoplasmic resistance is much lower than junctional resistance, the signal attenuation will occur mainly at these junctions. We further assume a uniform distribution of gap junctions along the cell surface. The signal attenuation per unit length normal to the slit stimulus will then be greater in the direction perpendicular to the streak resulting in receptive fields elongated along the streak.

The mechanism described above is based upon a network of L_1 units each of which has a common orientation. A random distribution of orientations of elongated cells would result in more equal mean distances between junctions along any two orthogonal directions. This type of anatomical arrangement is in fact seen in Golgi-impregnated L_1 -type horizontal cells in the peripheral retina. Here, our preliminary physiological findings indicate that L_1 -type horizontal cells have response eccentricities closer to 1.0 with no consistently observed preferred orientation along any particular axis.

In this paper, we have shown that the streak region of the turtle retina contains at least one population of cells with specialised properties; the axonal arborisations and the receptive fields of L_1 horizontal cells in this region are elongated in the direction of the streak. The functional significance of the specialisations in the streak region may be better understood from ganglion cell recordings in this region. However, such measurements have only been made in the visual streak of another species, the rabbit¹. Amongst rabbit ganglion cells in this region, orientation selective units were found which responded only to slits of light which were oriented either parallel or perpendicular to the streak. Although a full understanding of receptive fields at the ganglion cell level requires a greater knowledge of signal processing in the inner and outer plexiform layers, the oriented horizontal cell receptive fields described here may have significance for the development of orientation specificity in the visual streaks of species whose retinas contain this specialisation.

We thank Walter Stewart for Lucifer dye, and E. I. DuPont de Nemours & Co. (Inc.) for the Pontamine Fast Turquoise. We thank Drs K. I. Naka and J. Schnapf for reading the manuscript and Dr R. Nelson for useful discussions.

RICHARD A. NORMANN
HELGA KOLB
MENACHEM HANANI
EFREM PASINO
RICHARD HOLUB

Laboratory of Neurophysiology,
National Institute of Neurological and
Communicative Disorders and Stroke,
National Institutes of Health,
Bethesda, Maryland 20014

Received 11 December 1978; accepted 15 May 1979.

- Levick, W. R. *J. Physiol., Lond.* **188**, 285-307 (1967).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106-154 (1962).
- Henry, G. H., Dreher, B. & Bishop, P. O. *J. Neurophysiol.* **37**, 1394-1409 (1974).
- Schiller, P. M., Finlay, B. L. & Volman, S. F. *J. Neurophysiol.* **39**, 1320-1333 (1976).
- Brown, K. T. *Vision Res.* **9**, 1053-1062 (1969).
- Baylor, D. A., Fuortes, M. G. F. & O'Bryan, P. M. *J. Physiol., Lond.* **214**, 265-294 (1971).
- Simon, E. J. *J. Physiol., Lond.* **230**, 199-211 (1973).
- Saito, T., Miller, W. H. & Tomita, T. *Vision Res.* **14**, 119-123 (1974).
- Lamb, T. D. *J. Physiol., Lond.* **263**, 239-255 (1976).
- Murakami, M. & Shimoda, Y. *Brain Res.* **144**, 164-168 (1978).
- Stewart, W. *Cell* **14**, 741-759 (1978).
- Nelson, R., Lutzow, A. V., Kolb, H. & Gouras, P. *Science* **189**, 137-139 (1975).
- Leeper, H. *J. comp. Neurol.* **182**, 777-794 (1978).
- Naka, K. I. & Rushton, W. A. H. *J. Physiol., Lond.* **192**, 937-961 (1967).
- Kaneko, A. *J. Physiol., Lond.* **213**, 95-105 (1971).
- Marmarelis, P. Z. & Naka, K. I. *Biophys. J.* **12**, 1515-1532 (1972).
- Nelson, R. *J. comp. Neurol.* **172**, 109-135 (1977).
- Lasansky, A. & Vallerger, S. *J. Physiol., Lond.* **251**, 145-165 (1975).
- Yamada, E. & Ishikawa, T. *Cold Spring Harb. Symp. quant. Biol.* **30**, 383-392 (1965).
- Lasansky, A. *Invest. Ophthalmol.* **11**, 265-274 (1972).
- Lasansky, A. *Phil. Trans R. Soc. B* **265**, 471-489 (1973).
- Raviola, E. & Gilula, N. B. *J. Cell Biol.* **65**, 192-222 (1975).
- Kolb, H. *J. Neurocytol.* **6**, 131-153 (1977).

Involvement of cyclic GMP phosphodiesterase activator in an hereditary retinal degeneration

CYCLIC NUCLEOTIDES mediate many aspects of normal cellular metabolism¹; thus, degradation as well as synthesis of these intracellular mediators must be strictly regulated. Phosphodiesterase (PDE), the enzyme of cyclic nucleotide catabolism, is present in mammalian tissues in multiple forms, which differ in substrate specificity, kinetic characteristics and sub-cellular localisation². Moreover, a calcium-dependent protein activator (now called calmodulin) has been characterised that specifically activates at least one of the PDE types^{3,4} although other types of PDE are known to be activator independent^{5,6}. Thus, several mechanisms are present *in vivo* which allow strict control of PDE. A unique cyclic GMP-PDE is compartmentalised in the outer segments of retinal photoreceptor cells⁷⁻¹⁰; its activity is low in the dark-adapted state but increases dramatically on light adaptation¹⁰⁻¹². The resulting drop in cyclic GMP content could serve as a chemical 'signal' in the normal visual process^{13,14}. However, despite much investigation of various cyclic nucleotide systems, no definitive information has been obtained which clearly links a disorder of cyclic nucleotide metabolism with a disease process elsewhere than in retina^{15,16}. We have recently presented preliminary evidence that an abnormality in cyclic GMP metabolism could be present in the

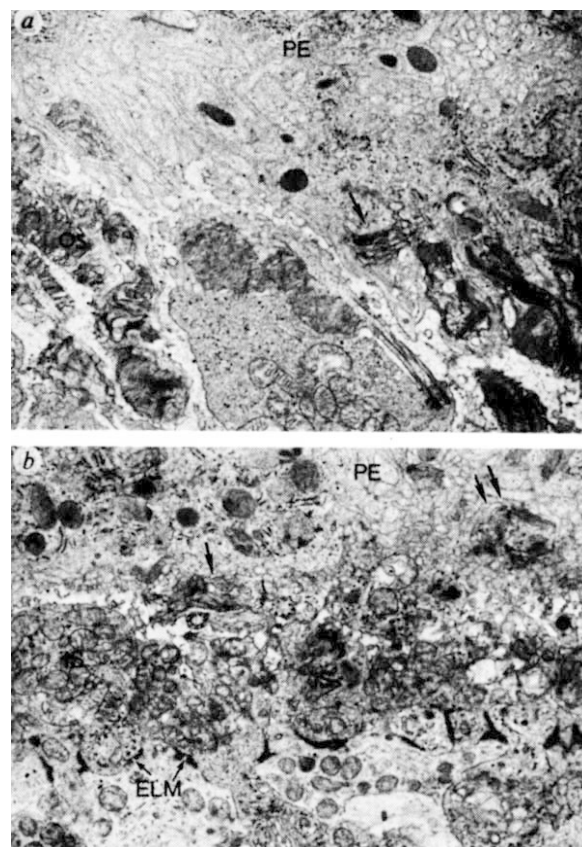


Fig. 1 *a*, Photoreceptor layer from a 13-d-old normal Irish setter. Photoreceptor outer segments have begun to elongate and contain well organised stacks of outer segment (OS) lamellar disks. Some of these seem to have been engulfed (arrow) by the pigment epithelium (PE). *b*, Photoreceptor layer from a 13-d old affected Irish setter. Small photoreceptor inner segments (IS) project through the external limiting membrane (ELM). There is minimal outer segment material; some is apposed to the PE apex (arrow) and some seems to be located within the PE cytoplasm (double arrow) $\times 10,400$.

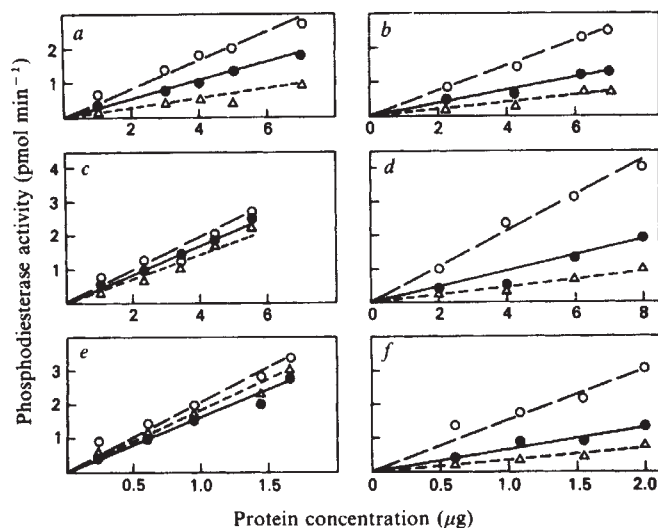


Fig. 2 Cyclic GMP phosphodiesterase activity in retinas of different ages. *a*, Control 9-d-old; *b*, affected 9-d-old; *c*, control 31-d-old; *d*, affected 34-d-old; *e*, control 48-d-old; *f*, affected 48-d-old. ●, No additions; ○, 10 μg purified brain activator added; △, 250 μM EGTA added. Experimental points are averages of triplicates which differed by no more than 15%.

retinas of Irish setter dogs with inherited rod-cone dysplasia¹⁵ that could lead to greatly increased cyclic GMP content, as had been reported in mice with inherited retinal degeneration¹⁶. We now report that the basic defect in the disease seems to be a failure to switch PDE type and a concomitant decrease in protein activator concentration early in postnatal development, at the time of photoreceptor differentiation.

The Irish setters used were bred to develop rod-cone dysplasia, and controls were heterozygous Irish setters unaffected by the disease. Tissues were obtained as previously described¹⁵; briefly, animals were anaesthetised with pentobarbital, their eyes enucleated and the retinas quickly dissected on ice and frozen in liquid nitrogen. Protein was determined by the method of Lowry *et al.*¹⁷ with bovine serum albumin as standard. Bovine brain activator and activator-deficient PDE were prepared as previously described¹⁸. The activator-deficient enzyme had a basal activity of 930 pmol cyclic AMP hydrolysed per mg protein per min; this was increased fourfold in the presence of optimal concentrations of Ca²⁺ and activator protein. PDE activity was measured using an assay mixture containing 40 mM Tris-HCl (pH 8.0), 3 mM MgSO₄, 50 μM CaCl₂, 1 μM ³H-cyclic AMP (36.6 Ci mmol⁻¹) or ³H-cyclic GMP (8.28 Ci mmol⁻¹) (NEN) and an appropriate amount of retinal homogenate in a total volume of 100 μl. When appropriate, 250 μM EGTA or 10 μg of purified brain activator was added before PDE assay. Samples were incubated for 12 or 15 min at 30°C, the reaction was terminated by boiling for 45 s, and 50 μg snake venom (*Crotalus atrox*, Sigma) was added. Samples were incubated for a second 10-min period at 30°C and 1.0 ml of IRP-58 resin (200–400 mesh, Rohm and Haas) added. After centrifugation, ³H-adenosine or ³H-guanosine in the supernatant was determined by liquid scintillation spectrometry. Recovery of ³H-nucleoside was >90%. Retinal protein activator was assayed by testing its ability to stimulate the activity of activator-deficient, purified cyclic nucleotide PDE enzyme (0.68 μg per 0.1 ml assay mixture) using boiled retinal aliquots as the source of activator. One unit of activator is defined as the amount required to give half-maximal stimulation of the activator-deficient PDE enzyme.

The second week after birth is critical in retinal development in dogs, as the photoreceptors begin to differentiate during this period (Fig. 1). At day 9 (Table 1 and Fig. 2*a, b*), cyclic GMP-PDE activity was already lower in the dystrophic dog (244 pmol per mg protein per min) than in heterozygous

controls (309 pmol per mg per min). However, cyclic AMP-PDE activity was the same in affected and control retinas at 9 days. Cyclic GMP-PDE activity in control retinas rose sevenfold (Table 1) by the time of photoreceptor maturation at 7 weeks¹⁹. Cyclic GMP-PDE activity observed in affected retinas during development was lower than that of control retinas, whereas no difference was observed in cyclic AMP-PDE activity (Table 1). Measurement of endogenous protein activator for cyclic GMP-PDE using activator-deficient brain PDE enzyme indicated lower amounts in affected than in control retinas in each age group studied. No differences were observed with the protein activator using ³H-cyclic AMP as substrate (not shown).

Activity of cyclic GMP-PDE was then investigated in the presence of exogenous brain activator protein or EGTA (Fig. 2). The latter was used because Ca²⁺ is necessary for activator-induced increases in PDE activity. The addition of brain activator increased PDE activity in the 9-d control retina but added EGTA decreased activity (Fig. 2*a*). In the 9-d affected retina (Fig. 2*b*), the inhibitory effect of EGTA was also apparent, as was the stimulatory effect of brain activator. At about 5 and 7 weeks, neither brain activator nor EGTA had any marked effect on PDE activity in the control retina (Fig. 2*c, e*). In contrast, EGTA continued to exert an inhibitory effect in the dystrophic retina, and the addition of brain activator restored PDE activity to the levels seen in the control (Fig. 2*d, f*).

The expression of PDE activity in cells seems complex in that it can be regulated by several factors including activator protein and Ca²⁺ (refs 20, 21) in addition to the basic activity of the enzyme itself. It has been suggested²² that the concentration of cyclic GMP in particular is controlled by a PDE protein activator. It was therefore not unexpected to find effects of activator and Ca²⁺ on cyclic GMP-PDE activity in 9-d control retinas. Morphologically, photoreceptor outer segments begin to develop at this time. However, it was surprising to find that after outer segment elongation and maturation added brain activator and Ca²⁺ had negligible effects on PDE activity in these retinas. This indicated that the activator-dependent PDE activity had switched to an activator-independent ('adult') type (Fig. 2). In contrast, in affected retinas, no such switch was found to occur as the outer segments began to develop. Exogenous activator markedly stimulated PDE activity throughout the 7-week experimental period, whereas endogenous activator levels fell to about half of that observed in control retinas. It is not known if endogenous activator levels are actually limiting in affected retinas. The finding that exogenous brain activator restores PDE activity to close to that seen in the control *in vitro*, however, could be important in future attempts to halt or at least slow the progress of the disease.

Table 1 Phosphodiesterase activity and protein activator activity in control and affected retinas

Animal age (day)	Animal type	Phosphodiesterase activity (pmol per mg protein per min)		Cyclic GMP-PDE activator concentration (units per mg protein)
		Cyclic AMP	Cyclic GMP	
9	Control	102	309	125
9	Affected	102	244	101
31	Control	104	887	53
34	Affected	104	326	23
48	Control	143	2,275	95
48	Affected	151	539	56

Phosphodiesterase activity was measured by determination of ³H-nucleoside with an ion-exchange method (see text). Protein activator was determined in boiled retinal aliquots using purified bovine brain PDE which was deficient in activator. One unit of protein activator activity is that amount which is required to give half-maximal stimulation of the activator-deficient brain PDE enzyme. Values given are averages of triplicate determinations in two experiments which agreed to within 15%.

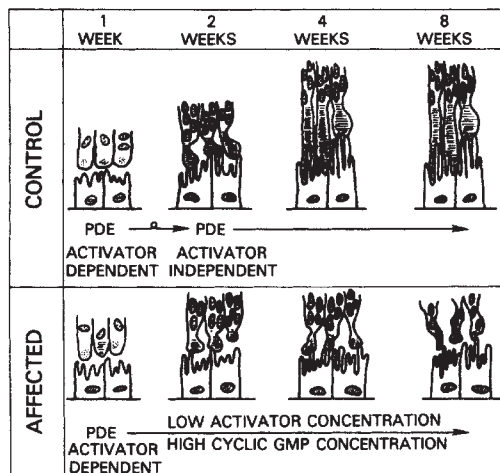


Fig. 3 Schematic outline of morphological and biochemical events in the development of outer segments (OS) in control and affected retinas. In both control and affected retinas, no OS are observed in the first postnatal week. OS appear in the second week, develop normally (4–5 weeks) in the control retinas but quickly degenerate in affected retinas. Biochemically, the cyclic GMP-PDE enzyme switches from an activator-dependent to an activator-independent type in control retinas but not in affected retinas. Protein activator concentration is substantially decreased in affected retinas.

It is well known that enzyme patterns change during development^{23,24}. Recent evidence suggests that, at least in some tissues, PDE activity changes during differentiation with regard to its sensitivity to regulation by the Ca^{2+} -dependent protein activator^{25,26}. Our data strongly indicate that a normal switch in PDE type fails to occur in affected retinas during photoreceptor development and that the low level of activator present in these retinas may not be sufficient to maintain adequate cyclic GMP-PDE activity (Fig. 3). It seems, therefore, that for the first time calmodulin can be directly linked to a disease process. It will be interesting to see if a similar situation exists in other degenerative diseases or if this biochemical defect is unique to inherited degeneration of the neural retina.

The dogs with rod-cone dysplasia were supplied by the National Retinitis Pigmentosa Foundation.

Y. P. LIU

National Cancer Institute,
National Institutes of Health

G. KRISHNA

National Heart, Lung and Blood Institute,
National Institutes of Health

G. AGUIRRE

Section of Ophthalmology,
University of Pennsylvania School of Veterinary Medicine,
Philadelphia, Pennsylvania

G. J. CHADER*

National Eye Institute,
National Institutes of Health,
Bethesda, Maryland 20014

Received 18 January; accepted 4 May 1979.

*To whom reprint requests should be addressed.

- Robinson, G., Butcher, R. & Sutherland, E. A. *Rev. Biochem.* **37**, 149–163 (1968).
- Appleman, M. & Terasaki, W. *Adv. Cyclic Nucleotide Res.* **5**, 153–162 (1975).
- Cheung, W. *Biochem. biophys. Res. Commun.* **29**, 478–482 (1967).
- Kakiuchi, S. & Yamazaki, R. *Biochem. biophys. Res. Commun.* **41**, 1104–1110 (1970).
- Uzunov, P. & Weiss, B. *Biochim. biophys. Acta* **284**, 220–226 (1972).
- Liu, Y. & Schwartz, H. *Biochim. biophys. Acta* **526**, 186–193 (1978).
- Chader, G., Johnson, M., Fletcher, R. & Bensinger, R. *J. Neurochem.* **22**, 93–99 (1974).
- Orr, H., Lowry, O., Cohen, A. & Ferrendelli, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4442–4445 (1976).
- Farber, D. & Lolley, R. *J. Neurochem.* **28**, 1089–1095 (1977).
- Bitensky, M. et al. *Adv. Cyclic Nucleotide Res.* **5**, 213–239 (1975).

- Chader, G., Herz, L. & Fletcher, R. *Biochim. biophys. Acta* **347**, 491–493 (1974).
- Yee, R. & Liebman, P. *J. biol. Chem.* **253**, 8902–8909 (1978).
- Goridis, C., Virmaux, N., Cailla, H. & Delaage, M. *FEBS Lett.* **49**, 167–169 (1974).
- Fletcher, R. & Chader, G. *Biochem. biophys. Res. Commun.* **70**, 1297–1302 (1976).
- Aguirre, G., Farber, D., Lolley, R., Fletcher, R. & Chader, G. *Science* **201**, 1133–1134 (1978).
- Farber, D. & Lolley, R. *Science* **186**, 449–451 (1974).
- Lowry, O., Rosebrough, N., Farr, A. & Randall, R. *J. biol. Chem.* **193**, 265–275 (1951).
- Lin, Y., Liu, Y. & Cheung, W. *J. biol. Chem.* **249**, 4943–4954 (1974).
- Hebel, R. *Adv. Anat. Embryol. Cell Biol.* **45**, 1–92 (1971).
- Teo, T. & Wang, J. *J. biol. Chem.* **248**, 5950–5955 (1973).
- Liu, Y. & Cheung, W. *J. biol. Chem.* **251**, 4193–4198 (1976).
- Kakiuchi, S., Yamazaki, R., Teshima, Y. & Uenishi, K. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3526–3531 (1973).
- Villee, C. *Fedn Proc.* **25**, 874–878 (1966).
- Greengard, O. & Thorndike, J. *Enzyme* **17**, 333–378 (1974).
- Tanigawa, Y., Shimoyama, M., Tai, J., Fujii, K. & Veda, I. *Biochem. biophys. Res. Commun.* **73**, 19–24 (1976).
- Singer, A., Dunn, A. & Appleman, M. *Archs Biochem. Biophys.* **187**, 406–413 (1978).

Evidence that cyclic GMP regulates membrane potential in rod photoreceptors

CYCLIC NUCLEOTIDES have been proposed as mediators of the decreased Na^+ permeability caused by illumination of vertebrate photoreceptors^{1,2}. We show here that cyclic GMP depolarises the rod outer segment (ROS) approximately to the Na^+ equilibrium potential within milliseconds after being injected intracellularly, that previous light adaptation antagonises this depolarisation, and that the injection of cyclic GMP without illumination initiates a repolarisation after a time lag which is longer than that following a light flash and proportional to the injection time. The results are interpreted in terms of a model in which cyclic GMP levels are controlled by the resultant of cyclase and phosphodiesterase (PDE) velocities. The PDE velocity of hydrolysis in the dark is increased by prior light adaptation in the presence of increased substrate (cyclic GMP). The data suggest that cyclic GMP is an important factor in the regulation of the membrane potential of the ROS.

Biochemical experiments on the cyclic nucleotide system of isolated homogenised ROS suggest that illumination decreases the concentration of cyclic GMP by activating a GTP-dependent PDE^{3,4}. The rate of light-initiated hydrolysis of cyclic GMP by PDE is apparently sufficiently fast to mediate the rapid initial decrease in Na^+ permeability^{5,6}. The cyclic GMP that is not hydrolysed presumably increases Na^+ permeability. The mechanism of increased Na^+ permeability of the ROS plasma membrane would be by phosphorylating a membrane protein^{6,7} as proposed for other tissues. Superfusion with cyclic GMP or a PDE inhibitor has been shown to depolarise rods⁸. We have injected cyclic GMP intracellularly through the recording pipette into ROS of the isolated retina of the toad, *Bufo marinus*, to determine the effects of excess cyclic GMP on membrane potential, latency and amplitude of the light response. We reported⁹ that this excess cyclic GMP increases both the latency and amplitude of the response. Here, using the same preparation and methods, we found that excess injected cyclic GMP depolarised the rod membrane. This depolarisation in dark-adapted preparations frequently exceeded the zero level (Figs 1, 2) and attained a maximum value of +11 mV (Fig. 1a, d). A reversal potential of 0 to +10 mV was obtained for the isolated *Necturus* rod¹⁰. A value of +11 mV implies an intracellular Na^+ concentration ($[\text{Na}^+]_i$) of 66 mM, based on our extracellular concentration of 103 mM. This compares with 50 mM $[\text{Na}^+]_i$ estimated using another technique¹¹. Therefore, +11 mV seems to approximate to the equilibrium potential for Na^+ . The maximum depolarisation we observe is consistent with the interpretation that the injected excess cyclic GMP renders the ROS plasma membrane permeable to Na^+ .

Intracellular injection of cyclic GMP in the absence of illumination should initiate a recovery from depolarisation after a time proportional to the amount injected, provided the dark concentration is less than the K_m of PDE which is 70 μM cyclic

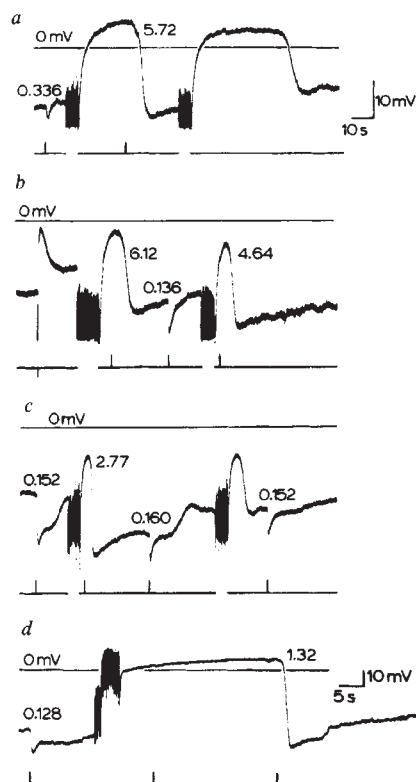


Fig. 1 Effects of intracellular injection of cyclic GMP in four different cells. Pen oscillations indicate injection of -2.5 nA; nominally 6×10^7 molecules cyclic GMP $\text{nA}^{-1} \text{s}^{-1}$ assuming transport number is 0.01. Numbers above membrane potential traces give latencies to light stimuli in seconds. All light flashes were 0.1 s, indicated by upward vertical lines on traces below each record. Light intensity is -3 log units on A and -2 log units for all other stimuli except the last on D, which is -1 log units. 0 log units = $5 \mu\text{W cm}^{-2}$. Change in noise level during injection in D may be due to a pipette tip which broke in the cell, releasing a large amount of cyclic GMP such that the first stimulus after injection failed to elicit a response. However, note small deflection that may originate from outside the rod compartment (which is about 500 rods by electrical coupling^{17,18}). Compartment size for cyclic GMP is not known.

GMP^{4,5}. There are $5\text{--}10 \times 10^7$ molecules of cyclic GMP per ROS in the dark¹², or a concentration of $40\text{--}80 \mu\text{M}$. The PDE should therefore be capable of higher velocities on injection of additional substrate. The V_{max} of PDE in the dark was estimated to be 10^4 times lower than in the light for frog ROS⁵. Nevertheless, a brief (40 ms) injection of cyclic GMP (Fig. 1b) causes an almost immediate depolarisation and a rapid repolarisation. Figure 2 shows that the time before repolarisation is proportional to the injection time. Each injection of additional substrate increases the PDE velocity, according to classical enzyme kinetics. The increased hydrolysis of cyclic GMP would tend to return the membrane potential to the baseline if cyclic GMP, acting through its kinase, determines membrane potential.

Following an intense flash (Fig. 3a) and after the baseline has returned to previous levels, injection of cyclic GMP fails to elicit

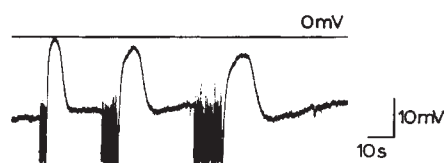


Fig. 2 Cell was dark adapted 10 min before injection; 2, 5 and 10 s injections of -2.5 nA, 25 mM cyclic GMP pipette.

a depolarisation. This suggests that light activation of PDE can potentiate the elevated velocity of PDE caused by additional substrate, thus preventing the depolarisation, or rapidly suppressing a brief depolarisation obscured by the pen oscillations. According to our model, the high V_{max} resulting from PDE light activation decays approximately with the time constant of dark adaptation (note no depolarisation after second injection, Fig. 3a, and small depolarisation 1 min later on, Fig. 3b). This persistently high V_{max} is unmasked by the injection of excess cyclic GMP in the dark.

Similar arguments may explain the length of latency, the increase in amplitude and the steeper repolarisation following a light flash delivered during the depolarisation after an injection of cyclic GMP. The light flash would briefly activate PDE, but the persistently high velocities of PDE caused by prior illumination and high substrate concentrations would be necessary for rapid hyperpolarisation of the membrane potential below the baseline. Illumination initiates hydrolysis by activation of PDE, but whereas the control hyperpolarisation in Fig 1a is complete in less than 0.5 s, the light-initiated PDE activity must have been augmented by increased substrate plus the light adaptation

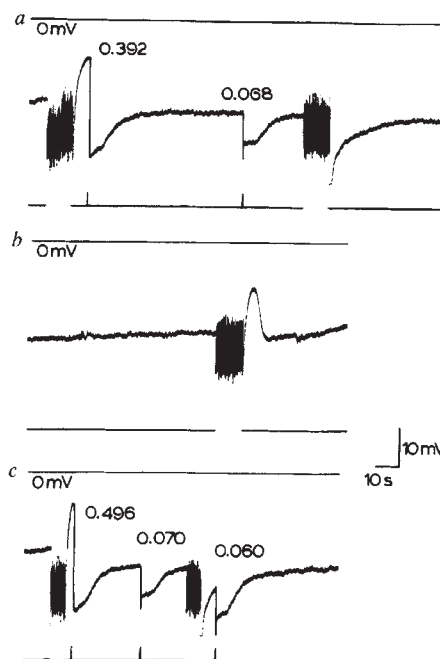


Fig. 3 a, b And c continuous recordings. Injections as in Figs 1 and 2, 25 mM cyclic GMP pipette. All light stimuli 0 log units.

effect mentioned above during the 5.72 s in which there is no visible response. The combination of increased substrate and light also seems to cause a persistent increase in PDE velocity, or undershoot, even after cyclic GMP levels are reduced, as judged by the increased plateau phase (Fig. 1c).

Figure 4 shows a control experiment in which the injection of the product of cyclic GMP hydrolysis, 5'-GMP, causes a small depolarisation. The response to illumination during this depolarisation is smaller than the control and of the same latency, in contrast to the opposite effects caused by cyclic GMP.

Excitation and adaptation of photoreceptors by light is a complex process that may involve molecules of the cyclic nucleotide systems, Ca^{2+} (refs 13–16) and other molecules. The experiments reported here are consistent with the interpretation that excess cyclic GMP injected into ROSs is capable of

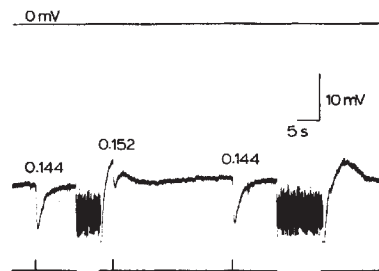


Fig. 4 Injections of -2.5 nA, 25 mM $5'$ -GMP pipette. Light stimuli -2 log units.

depolarising the rod plasma membrane to the Na^+ equilibrium potential, that both the injection and the previous adaptational history affect the intracellular concentration of cyclic GMP through the hydrolysing capacity of PDE, and that cyclic GMP must be hydrolysed by light-activated PDE so as to hyperpolarise the membrane to produce the normal response to illumination. These experiments suggest that normal intracellular levels of cyclic GMP may be an important determinant of the latency and other response characteristics of vertebrate rod photoreceptors.

We thank Ted Reid for helpful discussions. This work was supported in part by USPHS grant EY00089.

Note added in proof: Control experiments indicate that the zero level shown in the figures should be shifted upwards by 5 mV because of differences in tip potentials measured in Conway's and mock intracellular superfusates based on ref. 11.

WILLIAM H. MILLER
GRANT D. NICOL

Yale University School of Medicine,
Departments of Ophthalmology and
Visual Science, and Physiology,
New Haven, Connecticut 06510

Received 17 April; accepted 18 May 1979.

1. Bitensky, M. W., Gorman, R. E. & Miller, W. H. *Proc. natn. Acad. Sci. U.S.A.* **68**, 561–562 (1971).
2. Miller, W. H., Gorman, R. E. & Bitensky, M. W. *Science* **174**, 295–297 (1971).
3. Chader, G. J., Bensinger, R., Johnson, M. & Fletcher, R. T. *Expl Eye Res.* **17**, 483–486 (1973).
4. Miki, N., Keirns, J. J., Marcus, F. R., Freeman, J. & Bitensky, M. W. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3820–3824 (1973).
5. Yee, R. & Liebman, P. A. *J. biol. Chem.* **253**, 8902–8909 (1978).
6. Bownds, D. *et al. Soc. Neurosci. 8th A. Meet. St Louis, abstr.* (1978).
7. Lolley, R. N., Brown, B. M. & Farber, D. B. *Biochem. biophys. Res. Commun.* **78**, 572–578 (1977).
8. Lipton, S. A., Rasmussen, H. & Dowling, J. E. *J. gen. Physiol.* **70**, 771–791 (1977).
9. Nicol, G. D. & Miller, W. H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5217–5220 (1978).
10. Bader, C. R., MacLeish, P. R. & Schwartz, E. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3507–3511 (1978).
11. Normann, R. A. & Perlman, I., *Invest. Ophthalm. visual Sci., ARVO Abstr.* 219 (1978).
12. Woodruff, M. L., Bownds, D., Green, S. H., Morrissey, J. L. & Shedlovsky, A. *J. gen. Physiol.* **69**, 667–679 (1977).
13. Yoshikami, S. & Hagins, W. A. *Biophys. J.* **10**, 60a (1970).
14. Hagins, W. A. & Yoshikami, S. in *Vertebrate Photoreception* (eds Barlow, H. B. & Fatt, P.) 97–139 (Academic, New York, 1977).
15. Brown, J. E., Coles, J. A. & Pinto, L. H. *J. Physiol., Lond.* **269**, 707–722 (1977).
16. Bertrand, D., Fortes, M. G. F. & Pochobradsky, J. *J. Physiol., Lond.* **275**, 419–437 (1978).
17. Leeper, H. F., Normann, R. A. & Copenhagen, D. R. *Nature* **275**, 234–236 (1978).
18. Gold, G. H. *J. Neurophysiol.* **42**, 311–327 (1979).

Opposite effects of intracellular Ca^{2+} and glucose on K^+ permeability of pancreatic islet cells

MANY cells possess a potassium permeability system that is activated by an increase in the concentration of Ca^{2+} in the cytosol. This Ca^{2+} -sensitive permeability was originally described¹ and is best characterised² in erythrocytes, where its function is still not known. However, a similar mechanism could represent a physiological control of the membrane conductance in excitable tissues such as nerve^{3,4} or cardiac muscle⁵. Physiological stimulation of salivary glands results in a Ca^{2+} -medi-

ated release of potassium⁶ and hyperpolarisation⁷. In contrast, glucose depolarises pancreatic B cells^{8,9} by decreasing their potassium permeability^{10,11}, but also stimulates Ca^{2+} uptake by islet cells^{12–14}. I report here an investigation into the possible interplay between these two effects of glucose. Evidence that intracellular Ca^{2+} increases the K^+ permeability in pancreatic islet cells and that the glucose-stimulated Ca^{2+} inflow may represent a feedback control of the glucose-mediated decrease in K^+ permeability of B cells is presented.

The K^+ permeability of pancreatic islet cells was evaluated by continuously monitoring the efflux of $^{86}\text{Rb}^+$ from preloaded isolated rat islets in a perfusion system^{10,14,15}. The simultaneous measurement of $^{86}\text{Rb}^+$ and $^{42}\text{K}^+$ loss from islets labelled with both tracers has shown that the efflux of the more convenient isotope $^{86}\text{Rb}^+$ is a valid index of the qualitative changes in K^+ efflux from islet cells (data not shown).

A sudden rise in the extracellular Ca^{2+} concentration (0.5 to 10 mM) led to a reversible increase in the rate of $^{86}\text{Rb}^+$ efflux from islet cells perfused with a medium containing 3 or 6 mM

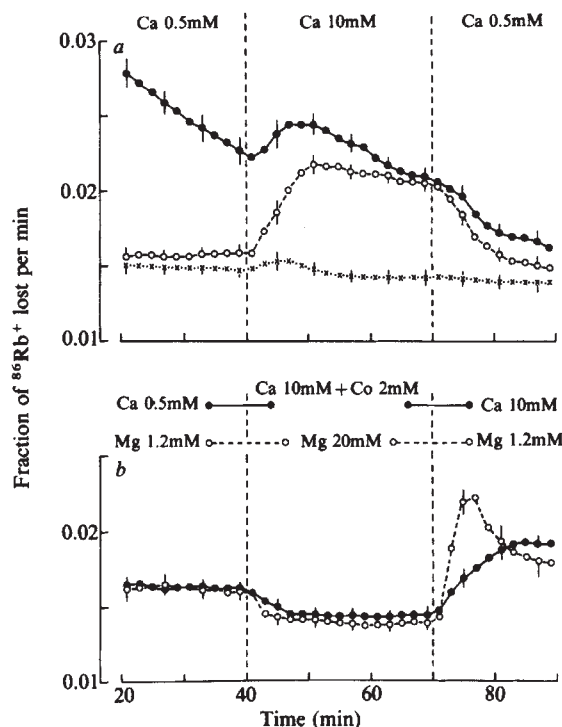


Fig. 1 Effect of a high concentration of extracellular calcium on rubidium efflux from perfused rat islets. *a*, Ca^{2+} concentration was increased from 0.5 to 10 mM between 40 and 70 min; glucose concentrations were kept constant at: $\bullet$, 3 mM; $\circ$, 6 mM; $\times$, 10 mM. *b*, 2 mM Co^{2+} was added to the medium containing 10 mM Ca^{2+} ($\bullet$), or the concentration of Ca^{2+} was kept constant (2.5 mM), while the concentration of Mg^{2+} was raised to 20 mM ($\circ$); the glucose concentration of the medium was constant at 6 mM. In (*a*) and (*b*), PO_4^{3-} was omitted to avoid precipitation of insoluble salts. Values are means $\pm$ s.d. of four experiments except in (*a*) where $n = 7$ at 6 mM glucose. Techniques have been detailed previously^{10,14,15} and are only outlined below. Isolated rat islets were obtained by collagenase digestion of the pancreas of fed male Wistar rats. All experiments were carried out at 37°C , with a Krebs-Ringer bicarbonate solution, pH 7.4, containing 6 mM K^+ and supplemented with 0.5% (w/v) bovine serum albumin. When Ca^{2+} was partially or completely omitted, it was replaced by Mg^{2+} ; when its concentration was increased, Na^+ was isosmotically decreased. Groups of 60 to 90 islets were first incubated for 150 min in 0.5 ml medium containing 3 mM glucose and 0.2 mM $^{86}\text{RbCl}$ (350–480 mCi mmol⁻¹, Radiochemical Centre). After three washings with non-radioactive medium, the islets were placed in perfusion chambers¹⁵. $^{86}\text{Rb}^+$ was counted by measuring the Cerenkov radiation¹⁰ in the effluent fractions, that were collected at 2 min intervals, and in the tissue at the end of the experiment. Results are expressed as fractional efflux of $^{86}\text{Rb}^+$.

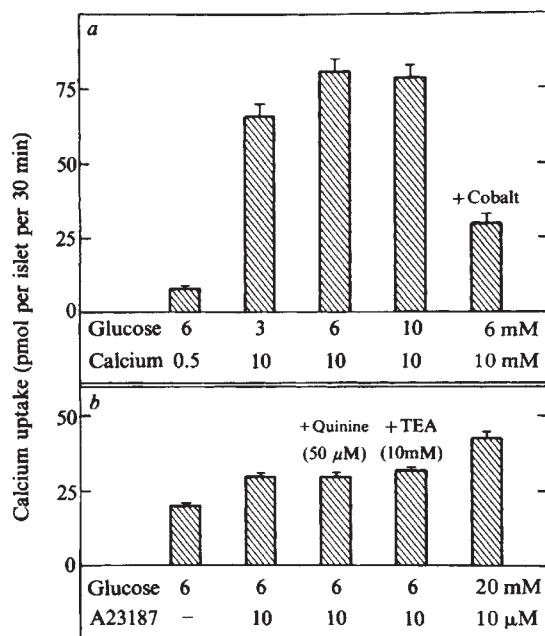


Fig. 2 Calcium uptake by islet cells. $^{45}\text{Ca}^{2+}$ uptake by rat islets was measured in relation to that of sucrose which is restricted to the extracellular space. The method has been described previously¹⁴. Islets were incubated in 100 μ l medium layered on silicone oil and separated from the radioactive solution by centrifugation through the oil at the end of the 30-min uptake period. The radioactivity entrapped in the tissue pellet was counted by liquid scintillation. Blanks without islets did not differ from the background. *a*, After preliminary incubation for 40 min in the presence of 0.5 mM CaCl_2 and the same concentration of glucose as for the incubation period (3, 6 or 10 mM), batches of 10 islets were transferred into the prewarmed incubation medium. This phosphate-free medium contained the indicated concentration of glucose, 0.25 mM $[6,6\text{-}^3\text{H}]\text{sucrose}$ (0.1 Ci mmol^{-1}) and either 0.5 mM $^{45}\text{CaCl}_2$ (60 mCi mmol^{-1}) or 10 mM $^{45}\text{CaCl}_2$ (3 mCi mmol^{-1}). In one series, 2 mM Co^{2+} was also present. Values are means $\pm$ s.e.m. of 14–17 batches of islets. *b*, After preliminary incubation for 40 min in the presence of 2.5 mM CaCl_2 and 6 mM glucose, batches of 10 islets were transferred into the incubation medium. The concentrations of A23187 and glucose were as indicated. The medium contained 2.5 mM $^{45}\text{CaCl}_2$ (12 mCi mmol^{-1}). Values are means $\pm$ s.e.m. of 12–16 batches of islets.

glucose, but not in the presence of 10 mM glucose (Fig. 1*a*). By contrast, increasing the Mg^{2+} concentration (1.2 to 20 mM) slightly depressed $^{86}\text{Rb}^+$ efflux (Fig. 1*b*), showing that the effect observed with Ca^{2+} is not due to damage of the cell membranes by a high level of divalent cations. The mere presence of 10 mM Ca^{2+} in the medium is not sufficient to increase the K^+ permeability of islet cells. Thus, $^{86}\text{Rb}^+$ efflux was not increased, but even somewhat decreased when the Ca^{2+} -antagonist¹⁴, cobalt, was added to the medium simultaneously with the rise in extracellular Ca^{2+} (Fig. 1*b*). Figure 2*a* shows the increase in Ca^{2+} uptake that occurs in islet cells when they are transferred from a medium with a low (0.5 mM) to a medium with a high (10 mM) concentration of Ca^{2+} . Using this experimental protocol, it is difficult to know exactly what proportion of Ca^{2+} has been adsorbed or has entered the cells. Note, however, that only cobalt, and not 10 mM glucose, reduced Ca^{2+} uptake, although both conditions prevented the increase in $^{86}\text{Rb}^+$ efflux that is otherwise produced by the rise in extracellular Ca^{2+} .

To evaluate the role of intracellular Ca^{2+} in the activation of $^{86}\text{Rb}^+$ efflux further, the effects of the divalent cationophore A23187 (ref. 16) were studied. Addition of 10 μ M A23187 to a medium containing Ca^{2+} steadily increased $^{86}\text{Rb}^+$ efflux in the presence of 6 mM glucose, but this effect was considerably reduced at 10 mM glucose (Fig. 3*a*). In the absence of extracellular Ca^{2+} , A23187 was almost ineffective, whereas the

subsequent introduction of Ca^{2+} into the medium was followed by a prompt and marked increase in $^{86}\text{Rb}^+$ efflux, which was completely reversible when both A23187 and Ca^{2+} were withdrawn (Fig. 3*a*). It is thus clear that the ionophore, which has a weak affinity for monovalent cations¹⁶, does not substantially transport Rb^+ ions from islet cells and that, again, Mg^{2+} cannot substitute for Ca^{2+} in activating K^+ permeability. The Ca^{2+} -activated K^+ permeability is specifically inhibited by quinine in the erythrocyte² and is sensitive to tetraethylammonium (TEA) in certain neurones³. Figure 3*b* shows that the increase in $^{86}\text{Rb}^+$ efflux brought about by A23187 + Ca^{2+} was reduced by TEA and stopped by quinine or 20 mM glucose. Only the effect of quinine was not reversible. In the presence of 6 mM glucose, A23187 produced a 50% increase ($P < 0.001$) in Ca^{2+} uptake by islet cells (Fig. 2*b*); this stimulation was unaffected by quinine or TEA and was increased by 20 mM glucose. Thus, the inhibition of Ca^{2+} -activated Rb^+ efflux by these agents is not due to a primary reduction of Ca^{2+} transport into the cells.

Since islet cells^{17,18}, as most cells, possess an energy-dependent system to sequester Ca^{2+} , metabolic inhibitors should cause cellular organelles to release stored Ca^{2+} and increase the levels of the cation in the cytosol. Figure 4*b* shows that the rate of $^{45}\text{Ca}^{2+}$ efflux from perfused islets was augmented after addition of the mitochondrial uncoupler, dicumylol, and that this effect was unaffected by TEA or quinine. Dicumylol simultaneously increased the rate of $^{86}\text{Rb}^+$ efflux (Fig. 4*a*); TEA partially and reversibly prevented the increase, whereas quinine suppressed it almost completely. Omission of extracellular Ca^{2+} did not alter

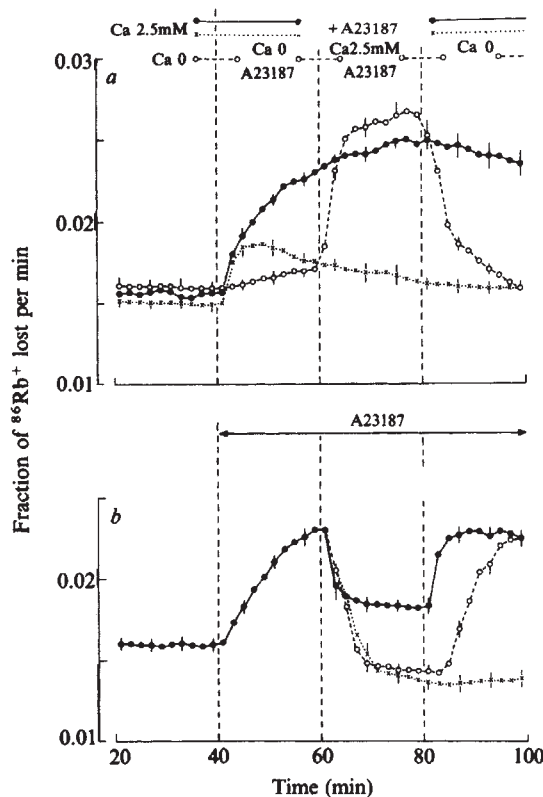


Fig. 3 Effect of Ca^{2+} -ionophore A23187 (10 μ M) on $^{86}\text{Rb}^+$ efflux from perfused rat islets. Experimental procedure as for Fig. 1. Glucose concentrations: *a*, $\bullet$ $\circ$, 6 mM; $\times$, 10 mM; *b*, 6 mM. A23187 (Lilly) was dissolved in dimethyl sulphoxide and the solvent was added alone (0.5%) in control solutions. In (*a*), A23187 was added from 40 to 100 min ($\bullet$ and $\times$) or 40 to 80 min ($\circ$); Ca^{2+} (2.5 mM) was present throughout or only between 60 and 80 min ($\circ$). Values are means $\pm$ s.d. of four experiments. In (*b*), Ca^{2+} was present throughout and A23187 from 40 to 100 min; between 60 and 80 min, TEA ($\bullet$, 10 mM) or quinine ($\times$, 50 μ M) was added, or the glucose concentration was increased from 6 to 20 mM ($\circ$). Values are means $\pm$ s.d. of nine experiments until 60 min and of three experiments thereafter.

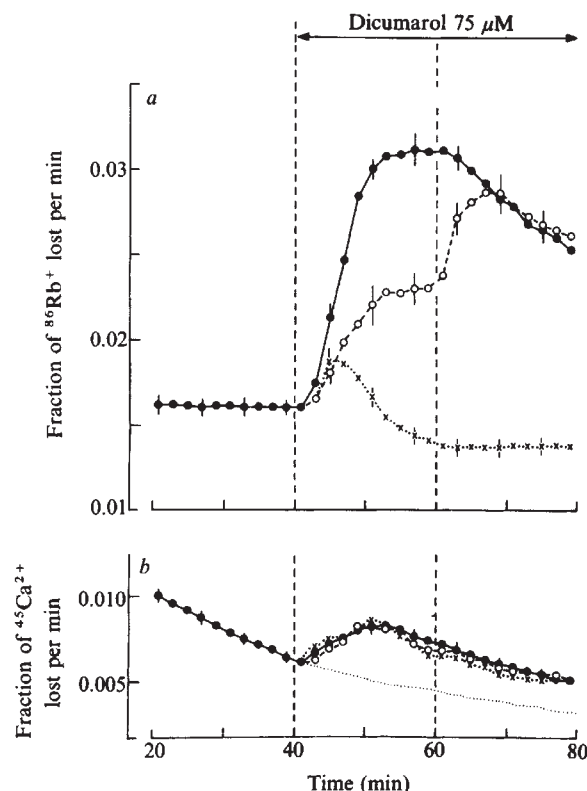


Fig. 4 Effect of dicumarol ($75 \mu\text{M}$) on $^{86}\text{Rb}^+$ (a) or $^{45}\text{Ca}^{2+}$ (b) efflux from perifused rat islets. Experimental procedure as for Fig. 1, except that for studying $^{45}\text{Ca}^{2+}$ efflux, islets were first incubated for 150 min in 0.5 ml medium with $1.25 \text{ mM } ^{45}\text{CaCl}_2$ ($200 \mu\text{Ci mmol}^{-1}$) and experiments were carried out in the absence of extracellular Ca^{2+} . Dicumarol (added from a 20 mM stock solution in 0.1 M NaOH) was present from 40 min onwards, while TEA ($\circ$, 10 mM) or quinine ($\times$, $50 \mu\text{M}$) was added between 40 and 60 min. The glucose concentration was constant at 6 mM . Values are means $\pm$ s.d. of nine experiments until 40 min and of three experiments thereafter.

the increase in $^{86}\text{Rb}^+$ efflux produced by dicumarol (not shown). Although a direct effect of dicumarol on the plasma membrane cannot be ruled out, these observations indicate that the activation of K^+ permeability by intracellular Ca^{2+} is not obligatorily linked to an influx of the divalent cation and are in keeping with the report⁸ that uncouplers of oxidative phosphorylation hyperpolarise B cells.

The experiments described above strongly suggest that a rise in intracellular Ca^{2+} increases the membrane permeability to potassium in islet cells as in other tissues¹⁻⁶. Although these experiments were made with whole islets comprising only two-thirds of B cells, the comparison of the changes in K^+ permeability with electrical recordings in B cells and with insulin release suggests physiological implications for our findings. When glucose is increased to a stimulatory concentration ($>5 \text{ mM}$), the membrane potential of B cells first decreases from the resting potential to a constant threshold level and then starts to oscillate in slow waves¹⁹. These waves are the regular alternation of depolarisation phases with bursts of spikes and silent periods of repolarisation. The duration of these phases of activity increases with the concentration of glucose and is correlated with insulin release¹⁹. It is also known that the depolarisation from the silent to the active membrane potential^{9,20,21} and perhaps also the spike activity²² depend on the entry of Ca^{2+} into the cells. However, high concentrations of extracellular Ca^{2+} shorten the periods of depolarisation²⁰ and decrease insulin release^{14,23}. On the other hand, TEA²⁴ and quinine²¹ suppress the polarisation phases of the slow waves, thus producing continuous activity, and markedly increase insulin release^{15,25}. In the light of these observations, this report

supports the suggestion that an activation of K^+ permeability by Ca^{2+} may be involved in the hyperpolarisation which terminates the bursts of activity. Within this framework, glucose-stimulated Ca^{2+} uptake might seem irreconcilable with the prolongation of the phases of depolarisation. However, the key observation is that high glucose levels reduce or even suppress the Ca^{2+} sensitivity of that K^+ permeability system. Lew and Ferreira²⁶ have shown that the Ca^{2+} sensitivity of the potassium gating mechanism depends on the metabolic state of the erythrocyte. As glucose has to be metabolised by islet cells to control their K^+ permeability¹⁰, the present system could represent a link between B cell metabolism and membrane permeability.

In conclusion, like the membrane of bursting neurones³, the B cell membrane possesses a Ca^{2+} -activated K^+ permeability. Its glucose sensitivity makes the system a sensitive control mechanism of the periodical oscillations of the membrane potential and insulin release.

I thank M. Nenquin and F. Mathot for assistance and Professors R. Casteels and A. E. Lambert for reading the manuscript. I am Chargé de Recherches of the FNRS, Brussels. This work was supported by grant 3.4509.75 from the FRSM, Brussels and a grant-in-aid from Hoechst-Belgium.

J. C. HENQUIN

Unité de Diabète et Croissance,
University of Louvain School of Medicine, UCL 54.74,
1200 Brussels, Belgium

Received 9 April; accepted 22 May 1979.

1. Gardos, G. *Acta physiol. hung.* **10**, 185-193 (1956).
2. Lew, V. L. & Beaugé, L. in *Membrane Transport in Biology* Vol. II (eds Giebisch, G., Tosteson, D. C. & Ussing, H. H.) 104-115 (Springer, Berlin, 1979).
3. Meech, R. W. A. *Rev. Biophys. Bioengng* **7**, 1-18 (1978).
4. Krnjevic, K. *Physiol. Rev.* **54**, 418-540 (1974).
5. Isenberg, G. *Nature* **253**, 273-274 (1975).
6. Selinger, Z., Batzli, S., Eimerl, S. & Schramm, M. *J. biol. Chem.* **248**, 369-372 (1973).
7. Petersen, O. H. *Physiol. Rev.* **56**, 535-577 (1976).
8. Dean, P. M. & Matthews, E. K. *J. Physiol., Lond.* **210**, 255-264 (1970).
9. Meissner, H. P. & Schmelz, H. *Pflügers Arch. ges. Physiol.* **351**, 195-206 (1974).
10. Henquin, J. C. *Nature* **271**, 271-273 (1978).
11. Sehlin, J. & Täljedal, I.-B. *Nature* **253**, 635-636 (1975).
12. Hellman, B., Sehlin, J. & Täljedal, I.-B. *Am. J. Physiol.* **221**, 1795-1801 (1971).
13. Malaisse-Lagae, F. & Malaisse, W. J. *Endocrinology* **88**, 72-80 (1971).
14. Henquin, J. C. & Lambert, A. E. *Am. J. Physiol.* **228**, 1669-1677 (1975).
15. Henquin, J. C. *Biochem. biophys. Res. Commun.* **77**, 551-556 (1977).
16. Pfeiffer, D. R., Reed, P. W. & Lardy, H. A. *Biochemistry* **13**, 4007-4014 (1974).
17. Sehlin, J. *Biochem. J.* **156**, 63-69 (1976).
18. Sugden, M. C. & Ashcroft, S. J. H. *Diabetologia* **15**, 173-180 (1978).
19. Meissner, H. P. *J. Physiol., Paris* **72**, 757-767 (1976).
20. Meissner, H. P. & Preissler, M. in *Treatment of Early Diabetes* (eds Camerini-Davalos, R. A. & Hanover, B.) (Plenum, New York, in the press).
21. Atwater, I. & Beigelman, P. M. *J. Physiol., Paris* **72**, 769-786 (1976).
22. Matthews, E. K. in *Hormones and Cell Regulation* (eds Dumont, J. & Nunez, J.) 57-76 (North-Holland, Amsterdam, 1977).
23. Milner, R. D. G. & Hales, C. N. *Diabetologia* **3**, 47-49 (1967).
24. Meissner, H. P. & Henquin, J. C. *Diabetes* **27**, Suppl. II, 457 (1978).
25. Henquin, J. C., Horemans, B., Nenquin, M., Verniers, J. & Lambert, A. E. *FEBS Lett.* **57**, 280-284 (1975).
26. Lew, V. L. & Ferreira, H. G. *Nature* **263**, 336-338 (1976).

Variability of cell cycle times measured *in vivo* in embryonic chick retina by continuous labelling with BUdR

IN this report, cell cycle times were measured in the embryonic chick retina by continuous labelling with 5-bromodeoxyuridine (BUdR) and the fluorescence plus Giemsa¹ (FPG) technique. The measurement of cell cycle times in embryonic chick tissues *in vivo* is preferably carried out by continuous rather than pulse labelling because a single injection of a nucleotide precursor into an egg can form a long-lasting pool which remains available for at least 24 h (ref. 2). Continuous labelling with BUdR provides information about the number of S phases a metaphase cell has encountered since the start of labelling whereas the previous continuous labelling procedure³ using tritiated thymidine indicates that a mitotic cell has encountered an S phase but does

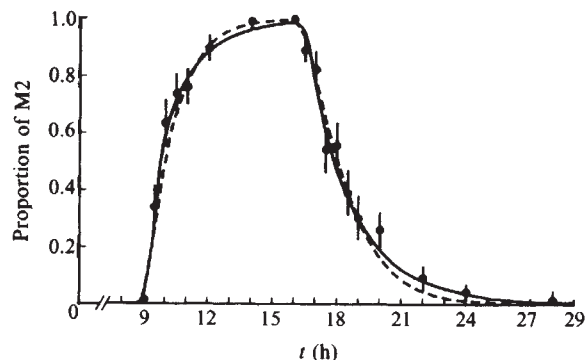


Fig. 1 The proportion of M2 at various times, t , after the start of labelling with BUdR. Each datum point (●) is the mean of five animals. Lines are drawn to 2 s.d. on either side of each datum point. 500 metaphase spreads were observed in each animal; the spreads which were not M2 at $t < 16$ h were either cells which had not incorporated BUdR or those which had incorporated it for one S phase; at $t \geq 16$ h, with very few exceptions they were cells which had incorporated BUdR for three or more S phases. To obtain these data, chick eggs incubated at 38 °C were injected at 72 h with 2×10^{-5} g BUdR and then sampled t hours later. Colchicine (1×10^{-5} g) was injected 1 h before sampling. Metaphase spreads were prepared from the neural retina as described previously². For FPG, the spreads were stained in 1×10^{-4} g ml⁻¹ 33258 Hoechst, mounted in 0.2 M sodium phosphate-citrate buffer, pH 7.2, and exposed to sunlight for 8–16 h, on successive days when necessary, and then stained in 3% Giemsa. The dose of BUdR was the lowest at which FPG differential staining could be achieved consistently. The continuous line is the curve of best fit obtained when the variances for phases SG₂EP and V were zero and when the phase G₁SAT (see text for definitions of these phases) was a constant plus a gamma distributed variable. The dashed line is the curve of best fit when G₁SAT was a constant plus an exponential variable. The curve for an inverse gaussian variable is not shown but it is close to that for the gamma.

not show how many. BUdR labelling thus gives more detailed data for estimating the means, variances and distributions of the durations of the cell cycle and its phases. Analysis of the data by statistical modelling has shown that the cell cycle is made up of a phase of constant duration and one of variable duration and we were able to test the appropriateness of different models for the variable phase. The variable phase was described well by a gamma distribution which was clearly not exponential in form and, somewhat less well, by an inverse gaussian distribution. A random transition in the cell cycle would be indicated by an exponentially distributed period; therefore, for the random transition model^{4,5} of the cell cycle to hold, there would have to be an exponentially distributed component within the variable phase. Our results do not exclude this but they do show that such a component could only be a minor part of the variable phase.

The procedure was to inject BUdR into eggs, sample the retina after various times and prepare² metaphase spreads. Colchicine was injected 1 h before sampling. The number of metaphase spreads which showed incorporation of BUdR for: one or no S phases, two S phases and three or more S phases, were counted in each sample. These three types of spread could be identified² by the FPG technique because it stains unsubstituted and unifilarly substituted chromatids darkly and bifilarly substituted chromatids lightly. The data are presented in Fig. 1 as the proportion of metaphases which had incorporated BUdR for two S phases (M2 metaphases) at various times after the start of labelling.

To describe the statistical analysis of the data, it is necessary to divide the cell cycle in a non-conventional way. A line is drawn at a point in S, a second line is drawn in prophase and a third at the end of metaphase. The line in S marks the beginning of the minimum time a cell must spend in S so as to incorporate enough BUdR to be detectable by FPG when the cell has undergone

another S phase. The line in prophase is at the beginning of the period during which the chromosomes could be rendered observable by the techniques for making metaphase spreads. This period is assumed to terminate at the end of metaphase. The new periods created by these boundaries were grouped as follows: G₁, the part of S following G₁ and the anaphase and telophase stages of mitosis were grouped together and called G₁SAT; the remainder of S, G₂ and a period approximating to early prophase were grouped and called SG₂EP; the period during which the chromosomes were observable was assumed to extend from about mid-prophase to the end of metaphase and was labelled V, for viewing period. For the analysis, the population of dividing retinal cells was modelled as an age-dependent branching process^{6,7} up to the time at which cells were acted upon by colchicine. The population was assumed to be growing exponentially⁷ from the time of BUdR injection and the parameters describing the population were assumed to remain constant from this time until the cells were collected. After colchicine began to act, it was assumed that all cells would be arrested in V. A parameter, c , known to be between 45 and 60 min (ref. 8), was introduced to denote the duration of colchicine action preceding t , the time of sampling. A theoretical curve for the proportion of V phase cells which were in M2 was derived as a function of t . This theoretical curve depended solely on c and the distribution of the durations of G₁SAT, SG₂EP and V and not on the individual distributions of G₁, S, G₂ and mitosis. This was why the phases of the cycle were grouped in the way described. Mathematical details of the analysis will be reported elsewhere⁹.

The theoretical curve was fitted to the data using normal, gamma, displaced gamma or inverse gaussian (reciprocal normal) distributions for G₁SAT, SG₂EP and V. The parameters for the best-fitting curves and the value of c were found by minimising a weighted sum of squares by computer search methods. The best fit in the initial trials always occurred when the variances of SG₂EP and V were zero. In subsequent searches, therefore, SG₂EP and V were modelled as constants and different distributions were tried for G₁SAT. Neither the

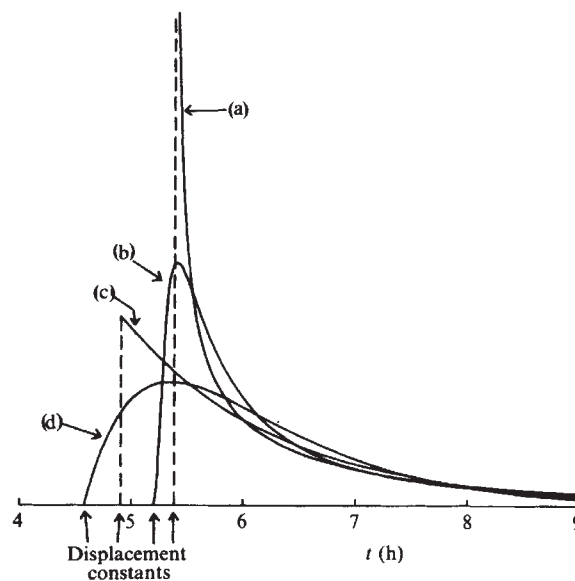


Fig. 2 Distributions for the variable phase in the cell cycle. Each curve is the probability density function obtained when G₁SAT was modelled by a displacement constant and a variable. The variables were: (a) a gamma with p unconstrained and estimated to be 0.44; (b) an inverse gaussian; (c) an exponential; (d) a gamma with p constrained to be 2; (a) gave the best-fitting theoretical curve and (b) was nearly as good; (c) was worse than either (a) or (b), and (d) gave a poorly fitting curve. The densities are markedly different for times less than 6 h both in shape and in the value of the displacement constant.

Table 1 Estimates of the cell cycle time

	Phase	Constant period (h)	<i>p</i>	Variable period (h)			Total period (h)	
				<i>b</i>	mean <i>p/b</i>	s.d. $p^{1/2}/b$	mean	median
Gamma	G ₁ SAT	5.4	0.44	0.34	1.3	2.0	6.7	5.9
	SG ₂ EP	1.7					1.7	1.7
	V	0.08					0.1	0.1
	Total cycle	7.2					8.5	7.7
Inverse gaussian	G ₁ SAT	5.2	0.48	0.30	1.6	2.3	6.8	6.0
	SG ₂ EP	1.7					1.7	1.7
	V	0.08					0.1	0.1
	Total cycle	7.0					8.6	7.8

The variable part of G₁SAT was modelled by a gamma density function

$$g(x) = b^p x^{p-1} \exp(-bx) / \Gamma(p) \quad (x > 0)$$

and an inverse gaussian density function

$$f(x) = \frac{p}{(2\pi b)^{1/2} x^{3/2}} \exp\{-\frac{1}{2}b(x - p/b)^2/x\} \quad (x > 0)$$

SG₂EP, V and the remainder of G₁SAT were constant. V is about 5 min (ref. 13); *c* = 0.75 h.

normal nor the gamma distributions alone gave good fits. The sum (convolution) of a normal distribution and an exponential distribution fitted best when the variance of the normal component was zero. This established that part of G₁SAT was also a constant. The fit was much improved when the variable part of G₁SAT was modelled by either the inverse gaussian or the gamma distribution instead of the exponential. The estimates for the cell cycle durations from the best-fitting curves are given in Table 1. The gamma distribution gave the best fit and for this the cell cycle time had a minimum of 7.2 h, a mean of 8.5 h and a standard deviation of 2.0 h.

These procedures show that SG₂EP, V and most of G₁SAT are constant in duration and constitute a constant phase in the cell cycle. The remainder of G₁SAT is variable and can be called the variable phase of the cell cycle.

The fit of the distributions assumed for the variable phase was assessed by the value of the minimised weighted sum of squares (WSS). The gamma distribution gave the lowest WSS and was thus the best fit. The inverse gaussian was nearly as good, with a WSS 1.24 times greater. The exponential distribution, on the other hand, gave a WSS 1.86 times that of the gamma. A more powerful test for the exponential distribution, because it is a special case of the gamma when the shape parameter *p* is equal to 1, was the value of *p* when a gamma distribution was assumed for the variable phase. We estimated *p* = 0.44 with an approximate standard error of 0.05 for the best-fitting theoretical curve. A value of *p* = 1 is thus approximately 11 standard deviations away from *p* = 0.44 and so it is very unlikely that our best-fitting curve had an exponentially distributed variable phase. The estimate of *p* shows that the total variability in cell cycle times cannot be attributed to a single random transition, nor can it be attributed to multiple transitions¹⁰ because they lead to a gamma distribution in which *p*, equal to the number of transitions, is an integer ≥ 2 and this is further from the estimate of *p*.

The existence of a random transition in the cell cycle depends on whether the total variability can be partitioned into an exponential distribution and some other distribution. Computations showed that a gamma distribution with *p* < 1 cannot be partitioned in this way. Our inverse gaussian distribution can have an exponential component but the mean of the exponential cannot sensibly be more than 20% of the mean of the variable phase. Therefore, either the exponential component does not exist or it is small, and a decision as to which of these two is so rests on the power to discriminate between the gamma and the inverse gaussian; this depends in particular on data about the very shortest cell cycle times (Fig. 2).

The advantage of our method is that it gives more information about the distributions of subdivisions in the cell cycle than do methods^{4,11} based on observations of total cell lifetimes. The latter methods have very little ability to distinguish between

alternative distributions for periods in the cell cycle.

A gamma distribution with *p* < 1 does not have a simple mechanistic explanation. An inverse gaussian distribution can have a simple interpretation such as the time taken to reach the critical cell size at which division occurs¹².

The model used for the analysis of the data assumes that the cells were in a steady state of exponential growth from the time of BUdR injection. The assumption is justified because (1) there are several thousand cells in a chick retina at 3 days incubation and so the population would have been asynchronous, and (2) independent observations by us (unpublished) have shown that the parameters describing the population did not change between 3 and 5 days incubation. These observations were the frequencies of unlabelled DNA strands in metaphase chromatids of chick retinal cells after continuous exposure to BUdR between 3 and 5 days incubation. The frequencies agreed well with those predicted by the model and hence show that the assumption of a steady state from the time of BUdR injection is valid.

We consider our observations to have been made on a homogeneous collection of cells and not on a mixture of subpopulations of cells cycling at different rates. The possibility that cell cycle times change in succeeding generations has been excluded by the evidence referred to for steady-state conditions and so the mixture of generations which is present in the population could not be a source of subpopulations cycling at different rates. However, there could be subpopulations which remain steady with time and which each have an exponentially distributed variable phase. This would be a way of explaining our findings in accordance with the random transition model of the cell cycle, for, although it is easy to prove mathematically that a mixture of exponentials does not lead to a gamma distribution, it might be difficult with the present data to discriminate between a gamma and the distribution resulting from a mixture of exponentials. Even so, we know of no evidence to suggest that such subpopulations exist in an embryonic chick retina.

We have presented evidence that the variability in cell cycle times is derived from one part of the cell cycle only and it is therefore in this part of the cycle that the rates of cellular proliferation would be controlled.

VALERIE B. MORRIS

School of Biological Sciences,
University of Sydney,
NSW 2006, Australia

RICHARD COWAN
DAVID CULPIN

CSIRO Division of Mathematics and Statistics,
PO Box 218 Lindfield,
NSW 2070, Australia

Received 6 March; accepted 22 May 1979.

1. Perry, P. & Wolff, S. *Nature* **251**, 156–158 (1974).
2. Morris, V. B. *Chromosoma* **60**, 139–145 (1977).
3. Puck, T. T. & Steffen, J. *Biophys. J.* **3**, 379–397 (1963).
4. Smith, J. A. & Martin, L. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1263–1267 (1973).
5. Shields, R. *Nature* **273**, 755–758 (1978).
6. Harris, T. E. *The Theory of Branching Processes* (Springer, Berlin, 1963).
7. MacDonald, P. D. M. in *The Mathematical Theory of the Dynamics of Biological Populations* (eds Bartlett, M. S. & Hiorns, R. W.) 303–314 (Academic, London, 1973).
8. Bloom, S. E. & Hsu, T. C. *Chromosoma* **51**, 261–267 (1975).
9. Cowan, R., Culpin, D. & Morris, V. B. (in preparation).
10. Takahashi, M. *J. theor. Biol.* **18**, 195–209 (1968).
11. Signoret, J. *J. Embryol. exp. Morph.* **42**, 5–14 (1977).
12. Wheals, A. E. *Nature* **267**, 647–648 (1977).
13. Hughes, A. F. W. & Fell, H. B. *Q. J. microsc. Sci.* **90**, 37–55 (1949).

Decreased response to calcitonin in osteopetrotic microphthalmic mouse bone

OSTEOPETROSIS is an autosomal recessive disease characterised by an increase in skeletal mass, often associated with retarded bone growth and a deficiency of marrow cavity development, which is related to a failure of bone resorption to keep pace with bone formation^{1,2}. The disease has been studied in man and in several genetic mouse and rat mutants¹. These mutants generally have diminished osteoclast function in the skeleton based on a lack of ruffled borders^{1,3}, lowered acid phosphatase activity⁴ and a lower level of parathormone-induced resorption *in vivo*^{3,5} and *in vitro*⁶. It seems that osteopetrosis has a cellular basis as it can be cured by parabiosis with a normal littermate⁷ or by administration of spleen or bone marrow cells⁸. However, the only report examining bone cell metabolism in osteopetrosis indicates that in the calvarium of the osteopetrotic incisor-absent rat, parathormone affects adenylate cyclase activity and lactic acid production to the same extent as in the bone from normal littermates⁹. Recently, specific hormone-sensitive biochemical activities in intact bone in culture^{10,11} or in cultured bone cells^{12,13} have been associated with the functions of osteoclasts and osteoblasts. Based on these biochemical parameters we show here that in osteopetrotic (microphthalmic) bone there is a defect in calcitonin response attributable to osteoclast function whereas osteoblast activity is apparently normal.

The animals used in all experiments were obtained by breeding our own C57BL/6J +/mi parental stock. Eye pigmentation pattern at birth was used to identify genotype. Calvaria were excised aseptically from 1–2-day-old homozygous dominant normals (+/+) and recessive (mi/mi) osteopetrotics and were cultured in minimum essential medium (GIBCO) supplemented with 10% fetal calf serum and the hormones indicated in Table 1. Air containing 5% CO₂ was the gas phase for all experiments except the resorption studies in which 50% oxygen, 5% CO₂ and 45% nitrogen¹⁰ was used. Cyclic AMP was measured by radioimmunoassay¹⁴ using the ¹²⁵I-tyrosine methylester of succinyl-cyclic AMP and rabbit anti-cyclic AMP antiserum (both from NEN). Parathormone had a biological activity >2,500 U mg⁻¹, and was isolated in our laboratory¹⁵.

The biochemical parameters chosen to test osteoclast function were the stimulation of cyclic AMP formation by both parathormone and calcitonin, and stimulation of hyaluronic acid synthesis by parathormone and the inhibition of this action by calcitonin. For osteoblast function the activities were stimulation of cyclic AMP formation and inhibition of citrate decarboxylation by parathormone—both effects are unaltered by calcitonin. To measure overall response of the osteopetrotic bone to parathormone net resorption of the tissue was measured.

Parathormone stimulated resorption in normal calvarium by more than 50% and calcitonin blocked this response (Table 1). In osteopetrotic calvarium basal resorption was similar to that in the normals but bone resorption was not stimulated by parathormone. The basal level of cyclic AMP in the normal and

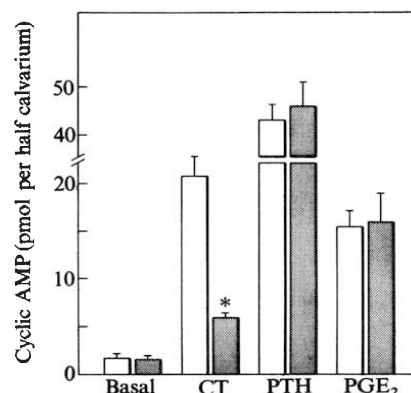


Fig. 1 The effect of calcitonin (CT, 12 nM), parathormone (PTH, 20 nM) and prostaglandin E₂ (PGE₂, 25 µM) on cyclic AMP formation in calvaria from normal and osteopetrotic mice. Half calvaria were placed in 1 ml of medium containing 0.5 mM theophylline and hormone additions as indicated. After 3 min cyclic AMP was extracted from the bones with 90% *n*-propanol²⁰ and was assayed by radioimmunoassay¹⁴. The dry weight and protein content of the normal and osteopetrotic were the same. Data are expressed as mean ± s.e.m. (*n* = 4). □, Normal (+/+); ■, osteopetrotic (mi/mi). * Significantly different from normal, *P* < 0.01.

osteopetrotic calvarium was the same (Fig. 1). Parathormone increased cyclic AMP production to the same extent in both groups as did prostaglandin E₂, another bone resorptive agent that apparently acts through cyclic AMP¹⁶. In contrast, calcitonin increased cyclic AMP more than 10-fold in the normal bones but only threefold in the osteopetrotic calvarium. Basal and parathormone-stimulated levels of hyaluronate synthesis were the same in both normal and osteopetrotic bones (Fig. 2). In normal bone calcitonin reduced this parathormone-stimulated increase by more than 75%. In osteopetrotic bone calcitonin exerted no inhibitory effect. Parathormone, as well as 1,25-dihydroxycholecalciferol, inhibited citrate metabolism in the osteopetrotic calvarium as effectively as in normal calvarium (Table 2).

These results show that osteopetrotic bone is deficient in certain biochemical markers that have previously been characterised in osteoclasts^{12,13}. With the demonstration that the tissue fails to resorb normally under the influence of parathormone the results are consistent with the morphological and cytological evidence¹ that the genetic lesion resides in osteoclasts. In contrast, osteoblasts of osteopetrotic bone responded normally to parathormone and other bone resorptive agents (prostaglandin E₂ and 1,25-dihydroxycholecalciferol). This observation,

Table 1 Effect of parathormone and calcitonin on calcium release from normal and osteopetrotic calvaria

Treatment	Ca release (µmol per bone per 48 h)	
	Normal	Osteopetrotic
Control	0.271 ± 0.011	0.305 ± 0.008
Parathormone	0.414 ± 0.025*	0.335 ± 0.015
Parathormone + calcitonin	0.309 ± 0.008	0.330 ± 0.022

Whole calvaria (parietal and frontal bones) were individually cultured in 1 ml minimum essential medium containing 10% fetal calf serum. Parathormone (20 nM) and synthetic salmon calcitonin (12 nM, Armour) were added as indicated. Calvaria were maintained at 37 °C in a gas phase of 5% CO₂, 50% O₂, 45% N₂ for 48 h (ref. 10). Medium was removed and calcium concentration determined¹⁹. The release of calcium is defined as the difference between live bone and dead bone (frozen and thawed three times) for each group. Values are mean ± s.e.m.; *n* = 7 (for normal) and *n* = 6 (for osteopetrotic).

* Different from control, *P* < 0.01.

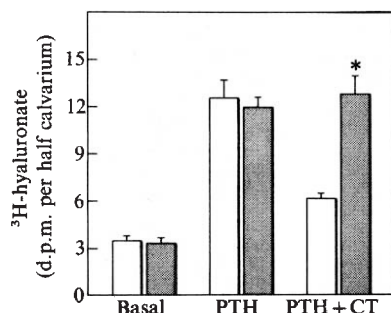


Fig. 2 Effect of parathormone (PTH, 20 nM) alone or in the presence of calcitonin (CT, 12 nM) on hyaluronate synthesis in half calvaria from normal and osteopetrotic mice. The bones were incubated for 24 h in medium (plus hormones as indicated) and then 3 μ Ci 3 H-glucosamine (NEN) was added to each bone. After 4 h the bones were rapidly frozen, digested with papain¹¹ and the 3 H-hyaluronic acid was isolated by differential cetylpyridinium chloride precipitation²¹. Data are expressed as mean $\pm$ s.e.m. ($n = 6$ for normal; $n = 7$ for osteopetrotic). Calcitonin alone had no effect on hyaluronate synthesis on either normal or osteopetrotic bone. $\square$, Normal; $\blacksquare$, osteopetrotic. * Significantly different from normal PTH+CT group, $P < 0.01$.

based on the degree of citrate decarboxylation^{12,17}, is consistent with that of Raisz *et al.*⁶ who found that parathormone reduced collagen synthesis by osteopetrotic and normal bone by the same amount.

A paradoxical aspect of these findings is that whereas the osteopetrotic bone failed to respond to parathormone in terms of bone resorption, it exhibited normal response in terms of cyclic AMP formation and hyaluronate synthesis. It is possible that despite the seemingly normal activation of adenylate cyclase by parathormone, the specific response of the osteoclasts were nonetheless impaired. It has been shown that parathormone elicits substantially less cyclic AMP per osteoclast than per osteoblast¹². This, together with the much larger number of osteoblasts than osteoclasts in calvarium¹, could have masked a decreased production of cyclic AMP by the osteoclasts. This is unlikely, however, as parathormone also increased the osteoclast marker hyaluronate synthesis to the same extent in osteopetrotic and normal bones.

The reduced response of osteopetrotic calvarium to calcitonin may be partly due to a decrease in the number of osteoclasts. Barnicot¹⁸ reported that there was about a 30% reduction in osteoclast number in the grey lethal osteopetrotic mutant. However, the fact that the levels of basal and parathormone-stimulated hyaluronate synthesis in the microphthalmic bone

were normal suggests that there was no difference in osteoclast number in our experimental conditions. It seems likely then that our results are better explained by a difference in the number or nature of calcitonin receptors, or an alteration in the mechanism for transmitting the calcitonin signal at the membrane level. The specific cause for such putative changes at the receptor level might be genetically based or secondary to alterations in circulating levels of calcitonin²² or other agents that can affect bone function. Examination of receptor binding in the intact calvarium and in cells isolated from the osteopetrotic bone would help elucidate the nature of the genetic lesion.

We thank James Stell for supervising and maintaining the colony of microphthalmic mice. Synthetic salmon calcitonin was a gift from Dr J. W. Bastian (Armour lot no. K713-801, 4700 U mg⁻¹) and 1,25-dihydroxycholecalciferol was a gift from Dr M. Uskokovic (Roche). This work was supported in part by grant DE05209 from the NIH.

G. NEIL KENT

Department of Biochemistry,
University of Kansas Medical Center,
Kansas City, Kansas 66103

DONALD G. WALKER

Department of Cell Biology and Anatomy,
The Johns Hopkins University School of Medicine,
Baltimore, Maryland 21205

DAVID V. COHN*

Calcium Research Laboratory,
Veterans Administration Medical Center,
Kansas City, Missouri 64128

Received 23 March; accepted 18 May 1979

* To whom correspondence should be addressed. Also in Department of Biochemistry, University of Kansas Medical Center.

1. Marks, Jr, S. C. & Walker, D. G. in *The Biochemistry and Physiology of Bone* Vol 4 (ed. Bourne, G.) 227-301 (Academic, New York, 1976).
2. Murphy, H. M. *Clin. Orthop.* **65**, 97-109 (1969).
3. Marks, Jr, S. C. *Am. J. Anat.* **138**, 165-190 (1973).
4. Walker, D. G. *Zellforsch. Mikrosk. Anat.* **72**, 100-124 (1966).
5. Marks, Jr, S. C. *Am. J. Anat.* **141**, 329-340 (1974).
6. Raisz, L. G., Simmons, H. A., Gworek, S. C. & Eilon, G. *J. exp. Med.* **145**, 857-865 (1977).
7. Walker, D. G. *Endocrinology* **91**, 916-920 (1972).
8. Walker, D. G. *Science* **190**, 784-785 (1975).
9. Rodan, G. A., Rodan, S. B. & Marks, Jr, S. C. *Endocrinology* **102**, 1501-1505 (1978).
10. Chu, L. L. H., MacGregor, R. R., Hamilton, J. W. & Cohn, D. V. *Endocrinology* **89**, 1425-1431 (1971).
11. Luben, R. A. & Cohn, D. V. *Endocrinology* **98**, 413-419 (1976).
12. Luben, R. A., Wong, G. L. & Cohn, D. V. *Endocrinology* **99**, 526-533 (1976).
13. Cohn, D. V. & Wong, G. L. in *Skeletal Research* (eds Simmons, D. & Kunin, A.) (Academic, New York, 1979) 3-20.
14. Steiner, A. L., Parker, C. W. & Kipnis, D. M. *J. biol. Chem.* **247**, 1106-1113 (1972).
15. Hamilton, J. W., Spier, F. W., MacGregor, R. R. & Cohn, D. V. *J. biol. Chem.* **246**, 3224-3233 (1971).
16. Klein, D. C. & Raisz, L. G. *Endocrinology* **86**, 1436-1440 (1970).
17. Wong, G. L., Luben, R. A. & Cohn, D. V. *Science* **197**, 663-665 (1977).
18. Barnicot, N. A. *Proc. R. Soc. B134*, 467-485 (1974).
19. Bett, I. M. & Fraser, G. P. *Clin. chim. Acta* **4**, 346-356 (1959).
20. Heersche, J. N. M., Fedak, S. A. & Aurbach, G. D. *J. biol. Chem.* **246**, 6770-6775 (1971).
21. Cleland, R. L. & Sherblom, A. P. *J. biol. Chem.* **252**, 420-421 (1977).
22. Murphy, H. M. *J. Endocr.* **53**, 139-150 (1972).

Table 2 Effect of parathormone, 1,25-dihydroxycholecalciferol and prostaglandin E₂ on citrate decarboxylation in normal and osteopetrotic calvaria

Treatment	Citrate decarboxylation (14 CO ₂ d.p.m. per half calvarium)	
	Normal	Osteopetrotic
Control	7,033 $\pm$ 1,383	7,305 $\pm$ 1,669
Parathormone	1,045 $\pm$ 114†	850 $\pm$ 107†
1,25-dihydroxycholecalciferol	1,675 $\pm$ 335†	2,091 $\pm$ 498*
Prostaglandin E ₂	1,091 $\pm$ 173†	1,168 $\pm$ 351†

Calvaria were halved along the sagittal suture and groups of 6 normal and 7 osteopetrotic half calvaria were individually cultured for each treatment group. Parathormone (20 nM), 1,25-dihydroxycholecalciferol (1 nM Roche) and prostaglandin E₂ (25 μ M, Sigma) were added as indicated. After 48 h of culture the capacity of the bones to decarboxylate [14 C]citrate (NEN) was measured as previously described¹⁰. Data are shown as mean $\pm$ s.e.m.

* Different from control, $P < 0.05$.

† Different from control, $P < 0.01$.

Melittin shares certain cellular effects with phorbol ester tumour promoters

MELITTIN, a 26-amino acid polypeptide of known sequence, is the major constituent of bee venom¹. Its primary amino acid sequence reveals that the molecule has both basic and hydrophobic termini, which give the molecule an amphipathic character. Consequently melittin can insert into phospholipid bilayers and exhibits surfactant activity. There is evidence that the association of melittin with cellular membranes results in: (1) disturbance of the acyl groups of phospholipids such that they

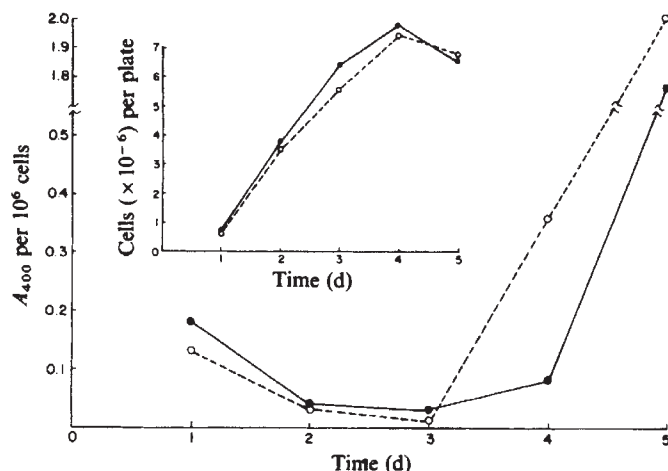


Fig. 1 The effect of melittin on the course of differentiation of C3B16 melanoma cells. The cells were plated in medium containing $2 \mu\text{g ml}^{-1}$ melittin in phosphate-buffered saline (PBS) or a control volume of PBS. After 24 h, fresh media containing either melittin or PBS were added. Cell counts and melanin determinations were carried out on replicate cultures every 24 h. The melanin content of cells plus medium was determined spectrophotometrically and is expressed in A_{400} units¹⁷. Cell number was determined with an electronic Coulter counter; inset shows total cell number per plate. All values are the means from three replicate culture plates; the s.e.m. was $\leq 10\%$ of the mean.

assume a broader range of angles with respect to each other^{2,3}; (2) increased phospholipid susceptibility to hydrolysis by phospholipase^{4,5}; and (3) increased synthesis of prostaglandins from the arachidonic acid released from the phospholipids⁶. Another amphipathic but structurally unrelated group of compounds consists of esters of the diterpene alcohol phorbol. These compounds are well known tumour promoters and have also recently been shown to have the ability to enhance phospholipid deacylation^{7,19}. It is therefore of interest to determine whether melittin can produce some of the other phenotypic effects that can be induced in cells in culture by phorbol ester tumour promoters^{8,9}. We report here that both melittin and the

phorbol ester, 12-O-tetradecanoyl phorbol-13-acetate (TPA), inhibit differentiation of mouse melanoma cells, enhance anchorage-independent growth of adenovirus-transformed rat embryo cells, and induce arachidonic acid release and prostaglandin synthesis from C3H/10T $\frac{1}{2}$ mouse embryo fibroblasts.

Changes in arachidonic acid metabolism were examined in normal C3H/10T $\frac{1}{2}$ mouse embryo fibroblasts as it is known that *in vitro* transformation by chemical carcinogens or radiation is enhanced by TPA and related phorbol esters^{10,11}. We found that TPA (2×10^{-8} M) and melittin (8×10^{-7} M) produced a two- to threefold enhancement of arachidonic acid and prostaglandin E₂ (PGE₂) release 3 h after addition to these cultures (Table 1). A low level of release of both arachidonic acid and PGE₂ in the control cultures was due to the presence of serum in the medium, an effect which has been previously reported^{7,19}. Both TPA and melittin, however, enhanced arachidonic acid and PGE₂ release even in serum-free conditions. Time-course studies indicated that the release of both arachidonic acid and PGE₂ was maximal 3 h after adding either TPA or melittin. After 3 h the levels declined and had returned to control values by 24 h.

The phorbol esters are very potent inhibitors of differentiation in several cell culture systems (see ref. 9, for review),

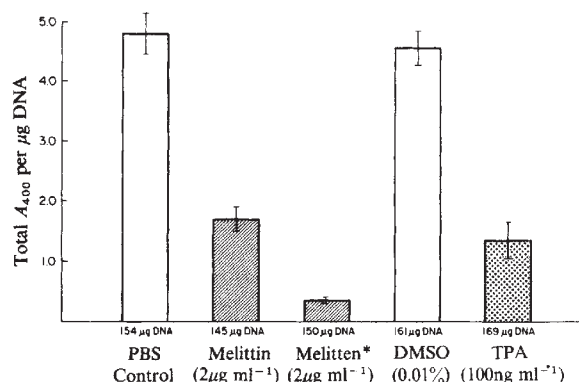


Fig. 2 The effect of melittin and TPA on the differentiation of C3B16 cells. The cells were plated in medium containing melittin, TPA or the appropriate vehicle controls. After 24 h the cells were changed to fresh medium containing the appropriate compounds, except in one group (*) in which the melittin concentration was increased to $5 \mu\text{g ml}^{-1}$. The cells and media were collected 5 d after plating. Melanin content of cells plus medium was determined spectrophotometrically¹⁷, and the DNA content of the cells measured by the diphenylamine method¹⁸. Melanin levels are presented as the mean $\pm$ s.e.m. of determinations from four replicate culture dishes. DNA values are total μg per plate on day 5.

Table 1 Changes in arachidonic acid and prostaglandin release from C3H/10T $\frac{1}{2}$ cells

Addition	Arachidonic acid (c.p.m. $\times 10^{-3}$ per 200 μl medium)	Prostaglandin E ₂ (c.p.m. $\times 10^{-1}$ per 200 μl medium)
Phosphate-buffered saline	16 ± 2.0	28 ± 5.3
Melittin (8×10^{-7} M)	32 ± 3.7	68 ± 7.1
DMSO 0.1%	17 ± 1.5	25 ± 6.0
TPA (2×10^{-8} M)	40 ± 5.1	98 ± 11.0

C3H/10T $\frac{1}{2}$ cells (2×10^3) were grown for one day in 50-mm plastic tissue culture dishes in 4 ml of Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Then fresh medium containing $2.5 \mu\text{Ci}$ of ^3H -arachidonic acid ($60\text{--}100 \text{ Ci mmol}^{-1}$; NEN) was added. After labelling for 24 h this medium was removed and the cells were rinsed three times with 5 ml of serum-free medium. The labelled cultures were then incubated with 2 ml of medium containing 10% fetal calf serum and melittin, TPA or the corresponding solvents. Aliquots of the media were collected at 3 h and analysed for arachidonic acid and PGE₂ content by thin layer chromatography, as previously described¹⁹. The values given are means $\pm$ s.e.m. from three replicate culture plates. DMSO, dimethyl sulphoxide. Additional studies indicated that during the labelling period 90% of the incorporated ^3H -arachidonic acid was in cellular phospholipids and that the trace amounts of unincorporated radioactivity associated with the cells could not account for the radioactivity subsequently released into medium.

including melanogenesis in mouse melanoma cultures^{12,20}. Cultures of the C3 clone of B16 mouse melanoma cells (C3B16) produce only small amounts of the pigment melanin until they reach confluence. There is then a marked and rather abrupt synthesis of melanin which accumulates in both the cells and culture medium.

Figure 1 indicates that when added to C3B16 cultures shortly after plating, melittin (8×10^{-7} M) caused a delay of about 24 h in the onset of melanogenesis. Nevertheless, the melittin-treated cultures grew at the same rate and reached confluence at the same time as the control culture (inset Fig. 1). Thus, the delay in differentiation caused by melittin was not due to growth inhibition or cytotoxicity. These effects are similar to those produced by TPA and other phorbol-related tumour promoters¹². Figure 2 compares in parallel the effects of melittin and TPA on pigment formation when either compound was added at the time the cells were plated. The amount of melanin was measured on day 5.

At 8×10^{-7} M ($2 \mu\text{g ml}^{-1}$) melittin caused a 66% inhibition of pigment formation. An approximately equivalent inhibition was

Table 2 Enhancement of growth in agar of E11 cells by TPA or melittin

Compound added	No. of colonies	Agar cloning efficiency (%)
Control	58 ± 7.8	2.9 ± 0.36
TPA (2×10^{-8} M)	192 ± 22.8	9.6 ± 1.14
Phorbol (3×10^{-7} M)	62 ± 8.4	3.1 ± 0.42
Melittin (8×10^{-7} M)	174 ± 16.8	8.7 ± 0.84

Agar cloning efficiency of E11 cells was determined as previously described¹⁵. Cells 2×10^3 were suspended in 4 ml of Dulbecco's medium (0.1 mM Ca^{2+} , and 0.4% Noble Agar) supplemented with 7.5% fetal bovine serum. This suspension was overlaid on prehardened 4-ml base layers of 0.8% Noble Agar in the same medium in 50-mm culture plates. Parallel cultures contained TPA, phorbol or melittin incorporated in both agar overlay and base layers. Cultures were refed after 4 and 10 d by overlaying with medium containing 0.4% Noble Agar and the appropriate compounds. Colony number was determined by light microscopy after 14 d (colony size >0.1 mm). Results are reported as the average number of colonies in 4 replicate plates, $\pm$ s.e.m. and the agar cloning efficiency (%), $\pm$ s.e.m. Neither the solvents PBS or 0.1% DMSO affected cloning efficiency.

obtained with 1×10^{-7} M TPA (100 ng ml^{-1}). Increasing the melittin concentration to 2×10^{-6} M ($5 \mu\text{g ml}^{-1}$) one day after seeding produced a further inhibition of melanogenesis (Fig. 2). In these studies the DNA content of cultures was measured and the data obtained (Fig. 2) also indicate the absence of any growth inhibition by either melittin or TPA at the concentration used.

Phorbol esters have also been shown to enhance the expression of transformation-related properties in various cell cultures^{8,9}. Growth of cells in soft agar (anchorage-independent growth) closely correlates with the tumorigenicity of cells *in vivo*¹³. Recent studies indicate that phorbol esters enhance the anchorage-independent growth of adenovirus-transformed rat embryo cells⁹ and certain serially passaged mouse epidermal cells¹⁴. We examined, therefore, the effect of melittin on anchorage-independent growth of E11 cells, a clone of adenovirus-transformed rat embryo cells^{15,16}. Table 2 indicates that both melittin (8×10^{-7} M) and TPA (2×10^{-8} M) caused approximately a threefold increase in the agar cloning efficiency of E11 cells, while phorbol, which is inactive as a mouse skin tumour promoter, did not enhance colony formation. Separate studies indicated that the concentrations of melittin and TPA used in these studies are approximately the optimal ones for these effects. Melittin concentrations greater than 2×10^{-6} M were cytotoxic.

The melittin preparation used in the above studies may contain trace amounts (~ 0.05 units per μg) of phospholipase A (Sigma) activity. This activity cannot, however, account for the results observed, as control studies using purified bee venom phospholipase A₂ (Sigma) in amounts several times higher than this level had no effect on melanogenesis by the B16 melanoma cultures, growth in agar of E11 cells, or PGE₂ synthesis.

Thus, it is rather striking that two structurally diverse compounds, melittin (a polypeptide) and TPA (a macrocyclic diterpene), have similar effects on three different properties of cells in culture: enhancement of the release of arachidonic acid and PGE₂; inhibition of melanogenesis; and enhancement of growth in agar. In view of the fact that melittin forms a very specific complex with the lipid phase of cell membranes¹⁻⁴, and the fact that a number of the early effects of TPA on cells are related to properties of cell membranes (see ref. 9 for review), it is likely that both compounds act by disturbing the structure and function of cell membranes. It is not clear, however, whether all the pleiotropic effects of the phorbol ester tumour promoters are secondary to such membrane effects or whether certain effects are mediated by non-membrane-related events. Melittin should prove highly useful for examining this question. In particular, it

will be of interest to determine whether melittin itself has tumour-promoting activity. These studies are in progress.

This research was supported by NCI contract No NO-1-CP2-3232 and in part by grant no. CA13696.

R. A. MUFSON
J. D. LASKIN
P. B. FISHER
I. B. WEINSTEIN

Division of Environmental Sciences and Cancer Center,
Institute of Cancer Research, Columbia University,
College of Physicians and Surgeons,
701 West 168th Street,
New York, New York 10032

Received 16 February; accepted 27 April 1979.

- Haberman, E. *Science* **177**, 314-322 (1972).
- Dawson, C. R., Drake, A. F., Helliwell, J. & Hider, R. C. *Biochim. biophys. Acta* **510**, 75-86 (1978).
- Verma, S. P., Wallach, D. F. & Smith, I. P. *Biochim. biophys. Acta* **345**, 129-140 (1974).
- Mollay, C., Kreil, G. & Berger, H. *Biochim. biophys. Acta* **426**, 317-324 (1976).
- Yunes, R., Goldhammer, A. R., Garner, W. K. & Cordes, E. H. *Archs Biochem. Biophys.* **183**, 105-112 (1977).
- Hassid, A. & Levine, L. *Res. Commun. chem. Path. Pharmac.* **18**, 507-517 (1977).
- Ohuchi, K. L. & Levine, L. *J. biol. Chem.* **253**, 4783-4790 (1978).
- Weinstein, I. B., Wigler, M. & Pietropaolo, C. *Origins of Human Cancer Book B* (eds Hiatt, H., Watson, J. & Winsten, J.) 751-772 (Cold Spring Harbor Laboratory, New York, 1977).
- Weinstein, I. B. *et al. Biological Markers of Neoplasia: Basic and Applied Aspects* (ed. Ruddle, W.) 451-471 (North-Holland, New York, 1978).
- Mondal, S., Brankow, D. W. & Heidelberger, C. *Cancer Res.* **35**, 2254-2260 (1976).
- Kennedy, A. R. & Little, J. B. *Nature* **276**, 825-826 (1978).
- Mufson, R. A., Fisher, P. B. & Weinstein, I. B. *Proc. Am. Ass. Cancer Res.* **19**, 183 (1978).
- Barrett, J. C. & Ts'o, P. O. P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3761-3765 (1978).
- Colburn, N. H., Former, B. & Warren, L. *Proc. Am. Ass. Cancer Res.* **19**, 65 (1978).
- Fisher, P. B., Goldstein, N. I. & Weinstein, I. B. *Cancer Res.* (in the press).
- Fisher, P. B., Weinstein, I. B., Eisenberg, D. & Ginsberg, H. S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2311-2314 (1978).
- Oikawa, A. & Nakayasu, M. *Yale J. Biol. Med.* **46**, 500-507 (1973).
- Burton, K. *Biochem. J.* **65**, 315-323 (1955).
- Mufson, R. A., DeFeo, D. & Weinstein, I. B. *Molec. Pharm.* (in the press).
- Mufson, R. A., Fisher, P. B. & Weinstein, I. B. *Cancer Res.* (in the press).

Quantitation of the adaptive response to alkylating agents

ESCHERICHIA COLI can acquire resistance to the lethal and mutagenic effects of alkylating agents such as *N*-methyl-*N'*-nitro-nitrosoguanidine (MNNG)¹. We have called this the adaptive response and have shown that it is distinct from previously described forms of DNA repair, including the inducible error-prone pathway^{2,3}. The induction of a similar response has recently been demonstrated independently in the liver cells of animals fed dimethylnitrosamine (DMN)⁴. The phenomenon is therefore apparently widespread and may conceivably protect us from mutagenesis and carcinogenesis induced by low levels of nitrosamines. The main mutagenic lesion produced by agents like MNNG and DMN is probably *O*⁶-methylguanine (*O*⁶-MeG)⁵, which can be read as adenine during DNA replication⁶ and transcription⁷; certainly, most of the mutations produced by MNNG are GC to AT transitions⁸. Therefore, it was not surprising to find that adapted bacteria contain less *O*⁶-MeG after a challenge dose of ³H-MNNG than do non-adapted bacteria⁹. However, the response seemed to have two distinct components, which we can call the fast and the slow reactions. Immediately after a 2-min exposure to ³H-MNNG the adapted bacteria contain less *O*⁶-MeG, as if some fast reaction were protecting them from accumulating alkylation at the *O*⁶ positions; there is, however, a limit to the capacity of this part of the response, and at higher concentrations of ³H-MNNG (or after longer periods of exposure) the adapted bacteria start to accumulate *O*⁶-MeG (and mutations) at the same rate as the non-adapted bacteria. This suggests that the molecules which carry out the fast reaction are used up during their interaction with *O*⁶-MeG (that is, each can function only once). During subsequent growth in the absence of MNNG, both adapted and

Table 1 The effect of temperature and chloramphenicol on the adaptive response

Challenge	Outgrowth	Non-adapted bacteria			Adapted bacteria		
		O^6 (c.p.m.)	N^7 (c.p.m.)	O^6/N^7	O^6 (c.p.m.)	N^7 (c.p.m.)	O^6/N^7
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	—	594	4,874	0.12	64	7,888	0.008
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 37 °C	14	3,550	0.004	—	—	—
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 37 °C, 100 $\mu\text{g ml}^{-1}$ CM	428	3,679	0.12	—	—	—
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG 100 $\mu\text{g ml}^{-1}$ CM	—	—	—	—	42	8,071	0.005
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	—	884	9,399	0.09	—	—	—
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 37 °C	0	7,907	0	—	—	—
10 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 0 °C	747	8,283	0.09	—	—	—
50 $\mu\text{g ml}^{-1}$ ^3H -MNNG	—	—	—	—	718	40,542	0.02
50 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 37 °C	—	—	—	0	21,010	0
50 $\mu\text{g ml}^{-1}$ ^3H -MNNG	60 min, 0 °C	—	—	—	346	20,054	0.02

Cultures of *E. coli* AB1157 were challenged for 2 min at 37 °C with ^3H -MNNG (56 mCi mMol $^{-1}$) and then filtered, resuspended and either processed immediately or allowed to undergo outgrowth and then processed as described previously⁹. In each case, the c.p.m. in DNA O^6 -methylguanine and N^7 -methylguanine are for samples of $\sim 3 \times 10^{10}$ bacteria. The adapted bacteria had been exposed, before challenge, to 0.5 $\mu\text{g ml}^{-1}$ MNNG for 90 min. For challenge of adapted bacteria in the presence of chloramphenicol (CM) the CM was added 10 min before challenge.

non-adapted bacteria lose their O^6 -MeG, but the process is completed much sooner in the adapted bacteria; thus, there is apparently a slow component to the adaptive response. This separation of the response into two components⁹ was supported by the observation that the fast reaction occurs at 0 °C whereas the slow reaction does not (P. F. Schendel and P.R., unpublished results). However, further experiments reported here indicate that these two reactions are different manifestations of a single, rather unusual form of repair.

Although adapted bacteria lose their O^6 -MeG during outgrowth sooner than non-adapted, this proved to be mainly because the fast reaction had ensured that they contained less O^6 -MeG at the start of the period of outgrowth. When the challenge doses of ^3H -MNNG were adjusted so that the adapted and non-adapted cultures were left with the same absolute amount of O^6 -MeG, the two cultures lost it at the same rate (Fig. 1). Thus, adapted bacteria seemed to differ from non-adapted solely in the presence of the fast reaction, and the slow reaction seemed to be some constitutive form of repair that could not be enhanced by adaptation.

Having reached that conclusion, we were therefore surprised to find that mutants, selected for their inability to adapt to MNNG, were unable to carry out either the fast or the slow reaction¹⁰. The explanation became obvious, however, when we remembered that the fast reaction is carried out by gene products that have accumulated during the period of adaptation and are apparently used up during the challenge. The slow reaction therefore could simply be the induced synthesis and immediate action of the gene products that carry out the fast reaction; following challenge, such synthesis would be occurring at the same rate in adapted and non-adapted bacteria. If this were correct, the slow reaction should be inhibited by chloramphenicol and should not occur at 0 °C (or in adaptationless mutants); in contrast, the fast reaction should not occur if chloramphenicol has been present during the period of adaptation¹, but it could conceivably occur at 0 °C and should not be inhibited by chloramphenicol present during the challenge. These predictions were borne out by the experiments shown in Table 1.

Far from being complicated, the adaptive response to O^6 -MeG seems therefore to be rather straightforward. If we accept that it is carried out by the product of the *ada* gene (in which we have isolated mutants)¹⁰ and assume that each of these molecules can dispose of only one O^6 -MeG, the numerology is approximately as follows. During growth in adapting concentrations of MNNG or during outgrowth following a challenge, each cell can make roughly 100 adaptation molecules per min, and so bacteria that divide every 30 min will contain about 3,000 molecules per cell when fully adapted. Any challenge that

creates less than 3,000 O^6 -MeG per cell can therefore be handled immediately by adapted bacteria, and no O^6 -MeG will be detectable at the end of the challenge. During subsequent outgrowth, all O^6 -MeG produced in non-adapted bacteria and any residual O^6 -MeG in adapted bacteria (that is, in excess of 3,000) will be handled at the rate of 100 per min. These numbers are derived from the published experiments using *E. coli* AB1157 (ref. 9). Other strains we have studied have a more effective response and seem to make about five times more of the gene product.

The most unusual feature of adaptation is that the molecules which mediate the response seem to be used up during the reaction. This could explain why large doses of alkylating agents will inhibit the capacity of cells to handle subsequent small doses^{9,11}. Furthermore, this feature may account for the discrepancy between the high carcinogenicity of certain nitrosamines and their low mutagenicity in the Ames test¹²; for reasons of economy, carcinogenicity is usually measured at fairly high dose rates at which the animal's adaptive response is

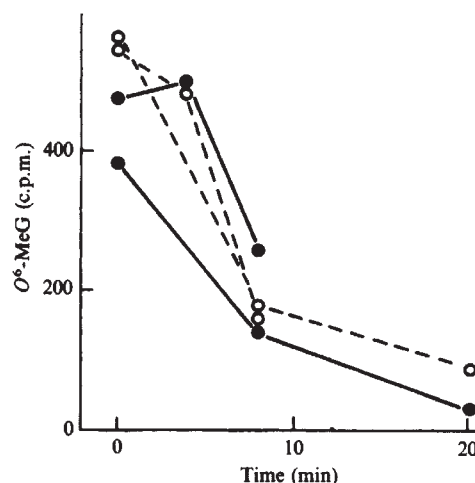


Fig. 1 The loss of O^6 -methylguanine from the DNA of adapted and non-adapted *E. coli* AB1157. A culture was adapted by growth for 90 min in the presence of 0.5 $\mu\text{g ml}^{-1}$ MNNG and then challenged with 50 $\mu\text{g ml}^{-1}$ ^3H -MNNG for 2 min; to produce approximately the same starting level of O^6 -MeG, the control non-adapted culture was challenged with 10 $\mu\text{g ml}^{-1}$ ^3H -MNNG. Both cultures were then filtered, resuspended and allowed to grow in the absence of MNNG. Samples containing 3×10^{10} bacteria were taken at various times and processed as described previously⁹. The results of two such experiments are shown. ●, Adapted bacteria; ○, non-adapted bacteria.

probably saturated, whereas bacterial mutagenicity is measured in growing cultures at low dose rates which are presumably within the capabilities of the bacterium's adaptive response.

Why the molecules that mediate adaptation can be used only once is not clear, but in this respect they resemble certain restriction enzymes which also can apparently act only once^{13,14}. Fortunately, it is now possible to study the exact chemistry of the interaction between the mediator of bacterial adaptation and its substrate, *O*⁶-MeG, because the process has been shown to occur *in vitro* (see accompanying paper¹⁵).

We thank Robert Kay for pointing out the similarity to restriction enzymes.

PETER ROBINS
JOHN CAIRNS

Imperial Cancer Research Fund,
Mill Hill Laboratories,
Burtonhole Lane,
London NW7, UK

Received 13 March; accepted 24 May 1979.

- Samson, L. & Cairns, J. *Nature* **267**, 281-283 (1977).
- Jeggo, P., Defais, M., Samson, L. & Schendel, P. *Molec. gen. Genet.* **157**, 1-9 (1977); **162**, 299-305 (1978).
- Schendel, P. F., Defais, M., Jeggo, P., Samson, L. & Cairns, J. *J. Bact.* **135**, 466-475 (1978).
- Montesano, R., Br  sil, H. & Margison, G. *Cancer Res.* **39**, 1798-1802 (1979).
- Loveless, A. *Nature* **223**, 206-207 (1969).
- Lawley, P. D. & Martin, C. N. *Biochem. J.* **145**, 85-91 (1975).
- Gerchman, L. L. & Ludlum, D. B. *Biochim. biophys. Acta* **308**, 310-316 (1973).
- Coulondre, C. & Miller, J. H. *J. molec. Biol.* **117**, 577-606 (1977).
- Schendel, P. F. & Robins, P. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6017-6020 (1978).
- Jeggo, P. (in preparation).
- Pegg, A. E. *Nature* **274**, 182-184 (1978).
- Meselson, M. & Russell, K. in *Origins of Human Cancer* (eds Hiatt, H. H., Watson, J. D. & Winsten, J. A.) 1473-1481 (Cold Spring Harbor, New York, 1977).
- Meselson, M., Yuan, R. & Heywood, J. A. *Rev. Biochem.* **41**, 447-466 (1972).
- Linn, S., Lautenberger, J. A., Eskin, B. & Lackey, D. *Fedn Proc.* **33**, 1128-1134 (1974).
- Karran, P., Lindahl, T. & Griffin, B. *Nature* **279**, 76-77 (1979).

Adaptive response to alkylating agents involves alteration *in situ* of *O*⁶-methylguanine residues in DNA

THE mutagenic lesion *O*⁶-methylguanine (*O*⁶-MeG) is slowly but actively removed from DNA in *Escherichia coli* cells treated with alkylating agents¹. In bacteria exposed to sublethal concentrations of chemical mutagens such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) or *N*-methyl-*N*-nitrosourea (MNUA), an apparently error-free repair system is induced that confers increased resistance to the mutagenic and lethal effects of challenge doses of the same or related compounds²⁻⁴. This adaptive response is associated with an increased ability of the induced *E. coli* to remove *O*⁶-MeG residues from their DNA (see ref. 5 and preceding paper⁶). We have investigated the action of cell-free extracts from adapted *E. coli* on DNA containing *O*⁶-MeG residues. Here we report that *O*⁶-MeG disappears from alkylated DNA after incubation with a crude enzyme fraction from adapted cells, although no concomitant release of the methyl group or the alkylated base or nucleotide occurs. Instead, this process is due to a previously unrecognised DNA repair mechanism involving enzyme-catalysed structural alteration of the alkylated residue.

The standard substrate used to study the removal of *O*⁶-MeG from DNA *in vitro* is *Micrococcus luteus* DNA (chosen because of its high guanine content) alkylated with [*Me*-³H]MNUA. Because the enzymatic release of free 3-methyladenine from the alkylated DNA by a DNA glycosylase⁷ complicates the analysis of the fate of *O*⁶-MeG residues, the MNUA-treated DNA was preincubated for 16 h at 80 °C in neutral solution. In this way, purines methylated in the 7 and 3 positions were released while the double-stranded DNA structure was maintained. After dialysis, about 75% of the radioactivity present in the DNA

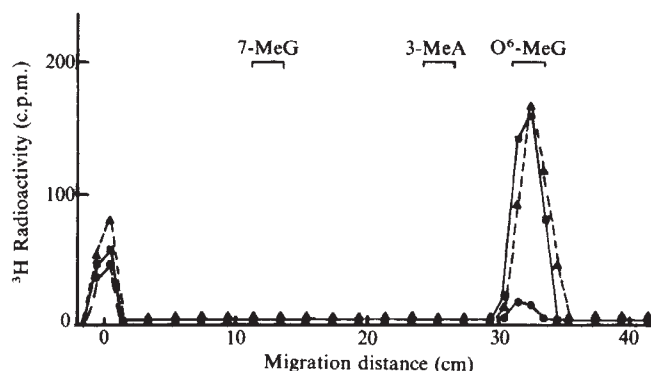


Fig. 1 Disappearance of *O*⁶-methylguanine from DNA catalysed by extracts of adapted *E. coli*. *M. luteus* DNA (25 mg) in 5 ml of 0.2 M sodium cacodylate and 1 mM EDTA (pH 7.2) was incubated for 2 h at 37 °C with 1 mCi ³H-MNUA (NEN, 1.03 Ci mmol⁻¹). The alkylated DNA was precipitated with 2 vols cold ethanol, dissolved in 5 ml of 0.1 M NaCl, 10 mM sodium citrate and 10 mM potassium phosphate (pH 7.4), and dialysed at 4 °C for 24 h against the same buffer. The DNA solution was then incubated at 80 °C for 16 h in a sealed glass ampoule, followed by dialysis at 4 °C for 24 h against 1 M NaCl, 10 mM Tris-HCl and 1 mM EDTA (pH 8.0), and for 4 h against the same buffer without NaCl. The DNA was stored frozen and had a specific radioactivity of 50,000 c.p.m. per mg. *E. coli* B/r was grown and adapted by exposure to MNNG (1 µg ml⁻¹ for 90 min) as described by Jeggo *et al.*². Adapted or non-adapted cells (3 g) were disrupted by grinding with three parts of sea sand and extracted with 12 ml of 70 mM HEPES-KOH, 1 mM dithiothreitol and 5% glycerol (pH 7.8). After removal of debris by centrifugation, extracts contained about 25 mg protein per ml. They were stored in small aliquots at -70 °C and were never refrozen. The reaction mixture (30 µl) contained 70 mM HEPES-KOH (pH 7.8), 1 mM dithiothreitol, 1 mM EDTA, 50 µM spermidine, 5% glycerol, 12 µg (600 c.p.m.) heated ³H-MNUA-treated DNA, and cell extract (2 µl). After 10 min at 37 °C, the reaction was stopped by chilling to 0 °C and addition of 30 µl cold 0.8 M trichloroacetic acid. After 10 min at 0 °C, the mixtures were centrifuged at 10,000g for 15 min. The supernatant solutions, which contained less than 5% of the total radioactivity, were discarded. The precipitates were suspended in 30 µl 0.1 M HCl, supplemented with carrier *O*⁶-methylguanine (30 µg in 3 µl H₂O), and the DNA was hydrolysed at 70 °C for 30 min. Each hydrolysate was applied to Whatman 3MM paper and chromatographed for 20 h in isopropanol/concentrated NH₃/H₂O (7:1:2). Reference methylated purines⁷ were run in separate lanes and localised as UV-absorbing material. Strips containing individual samples were cut out and then cut transversely into 1-cm pieces. Each such piece was cut into small fragments, which were transferred to a scintillation vial and eluted with 2 ml H₂O. Radioactivity was determined after addition of 15 ml of Triton X-based scintillation fluid. Reaction mixtures contained no extract (Δ), extract from non-adapted cells (■) or extract from adapted cells (●).

was in the form of *O*⁶-MeG residues (Fig. 1). The residual radioactive material may be ascribed to the presence of alkylated pyrimidine nucleotides and phosphotriesters⁸.

Incubation of this DNA substrate with crude cell extracts of *E. coli* B/r, adapted by growth for 90 min in the presence of 1 µg ml⁻¹ MNNG (ref. 2), results in the disappearance of *O*⁶-MeG residues from DNA. In the same conditions, extracts from non-adapted bacteria show no detectable activity. The method of assay involves incubation of the alkylated DNA with cell extract in an EDTA-containing reaction mixture, followed by precipitation of the DNA with cold ethanol or dilute trichloroacetic acid. After mild acidic hydrolysis of the precipitate to release free purines, paper chromatography of the hydrolysate is used to determine the amount of radioactivity co-migrating with authentic *O*⁶-MeG (Fig. 1). With increasing amounts of adapted cell extract in the reaction mixtures, decreasing amounts of *O*⁶-MeG are recovered (Fig. 2).

The method of assay clearly demonstrates the disappearance of *O*⁶-MeG from DNA, but an additional striking feature of the reaction is that no radioactive material is released in an ethanol- or acid-soluble form at the end of the incubation period. In conditions where 90% of the radioactive *O*⁶-MeG has been altered or removed, less than 5% of the total radioactivity is released in soluble form. On mild acidic hydrolysis of the precipitated DNA, using conditions allowing essentially quantitative release and recovery of *O*⁶-MeG (ref. 1), the radioactivity associated with the enzymatically altered residues is converted into volatile form. Thus, the radioactive material is

retained in the acid hydrolysate but is lost on evaporation or after application of the sample to a chromatography paper. Similar results have been obtained with DNA alkylated with either $[Me-^3H]MNUA$ or $[Me-^{14}C]MNUA$, and a volatile secondary product (presumably methanol) is generated by acid treatment in either case. These data indicate either that the methyl group bound at the O^6 position of guanine is transferred to another site by the activity in adapted *E. coli* cells, or that O^6 -MeG residues are destroyed (for example, by ring cleavage) without being released from DNA. Further, the enzymatically altered base residue is more acid labile than intact O^6 -MeG.

The radioactive material present in reaction mixtures treated with adapted cell extract (90% disappearance of O^6 -MeG from DNA) was not liberated in ethanol-soluble form (<10% release) by treatment with either Pronase (2 mg ml^{-1}) or pancreatic RNase ($100\text{ }\mu\text{g ml}^{-1}$) for 4 h at 30°C . On the other hand, treatment with pancreatic DNase I ($100\text{ }\mu\text{g ml}^{-1}$, 0.005 M each of $MgCl_2$ and $CaCl_2$ added) for 4 h at 30°C resulted in the conversion of more than 80% of the radioactivity into ethanol-soluble form. Similar results were obtained in control experiments with DNA incubated with non-adapted cell extract. Thus, the radioactive methyl groups in the ethanol-precipitable material generated from O^6 -MeG residues by treatment with an extract from adapted *E. coli* remain in DNA and do not seem to be transferred to a protein or RNA molecule.

These data differ from previous reports on the fate of O^6 -MeG residues in DNA after treatment with crude enzyme fractions. Kirtikar and Goldthwait⁹ claimed that a DNA glycosylase activity that would release free O^6 -MeG from alkylated DNA was present in a partly purified preparation of *E. coli* endonuclease II. However, an extensive search with different enzyme fractions from either non-adapted or adapted *E. coli* has failed to confirm the existence of such a DNA glycosylase (ref. 7 and our unpublished data). Pegg¹⁰ showed that rat liver extracts can remove O^6 -MeG from alkylated DNA, albeit about 1,000 times less effectively than the adapted *E. coli* extracts investigated here. The removal was not due to a DNA glycosylase, and the active enzyme was tentatively identified as a demethylase because part of the radioactivity from the substrate could be recovered as free methanol after prolonged incubation with the cell extract.

E. coli K-12 mutants defective in the adaptive response (*ada*) have recently been isolated¹¹. We have analysed extracts from two such mutants, *ada-5* and *ada-6*, and their *ada*⁺ parent strain AB 1157, for their ability to alter O^6 -MeG residues in DNA. Extracts ($10\text{--}40\text{ }\mu\text{g}$ protein added to the standard reaction mixture) from both mutants were unable to cause the

disappearance of detectable amounts (<10%) of O^6 -MeG from alkylated DNA even though the bacteria had been pretreated with MNNG, whereas the parent strain extract was clearly active. However, as *E. coli* K-12 yields a somewhat weaker adaptive response than *E. coli* B (refs 4, 6) the extract ($40\text{ }\mu\text{g}$ protein) from the *ada*⁺ strain only caused the loss of 60% of the O^6 -MeG residues. Growth rates in the conditions of adaptation were similar for both the parent and mutant strains. These data strongly indicate that the enzyme acting on O^6 -MeG residues investigated here is related to the previously characterised repair pathway of adaptation *in vivo*. The simplest interpretation of the available results is that this inducible enzyme is the product of the *ada*⁺ gene in *E. coli*.

There has been only preliminary characterisation of the properties of the *E. coli* enzyme activity associated with the adaptive response. The activity remains in the supernatant of crude cell extracts after precipitation of nucleic acids (and some proteins) with 0.8% streptomycin sulphate, and it can be recovered by precipitation between 1.6 M and 2.6 M ammonium sulphate followed by dialysis. It is heat labile (50% inactivation occurring after 1 min at 45°C in the standard reaction mixture without DNA) and is partly inactivated by freezing and thawing. There is no apparent cofactor requirement of the ammonium sulphate-fractionated enzyme, and no detectable stimulation of activity by Mg^{2+} or ATP. Inclusion of spermidine (5×10^{-5} to $2 \times 10^{-4}\text{ M}$) in the assay mixture results in slight (about twofold) stimulation of activity.

Although the present results show that O^6 -MeG removal in adapted *E. coli* does not occur by nuclease, glycosylase or demethylase action, the precise nature of the enzymatic alteration of this alkylated residue remains unclear. The methyl group might be transferred to another, biologically less hazardous, site in DNA. However, the present data exclude methyl group transfer to the N^7 position of guanine, because no 7-methylguanine or any other acid-stable product is generated (Fig. 1). Alternatively, the miscoding O^6 -MeG might be converted to a non-coding residue, such as by cleavage of the pyrimidine ring to generate a substituted imidazole residue remaining bound to DNA. Methyl-group transfer or ring-opening of the alkylated residue might also result in a lesion more readily recognised by other DNA repair enzymes. It is now of interest to define the structural alteration of O^6 -MeG that occurs in adapted *E. coli*, and this will depend on the availability of suitable methods to release and purify the labile altered residues from enzyme-treated alkylated DNA.

We thank John Cairns for helpful discussions, Penelope Jeggo for providing us with *E. coli ada* mutants, and Anita Jacobsson for technical assistance. This work was supported by the Swedish MRC. P.K. was the recipient of an EMBO fellowship (1977–79).

PETER KARRAN*
TOMAS LINDAHL

Department of Medical Chemistry,
University of Gothenburg,
400 33 Gothenburg, Sweden

BEVERLY GRIFFIN

Imperial Cancer Research Fund Laboratories,
Lincoln's Inn Fields,
London WC2, UK

Received 16 March; accepted 14 May 1979.

*Present address: MRC Cell Mutation Unit, University of Sussex, Falmer, UK.

1. Lawley, P. D. & Orr, D. J. *Chem.-biol. Interact.* **2**, 154–157 (1970).
2. Samson, L. & Cairns, J. *Nature* **267**, 281–283 (1977).
3. Jeggo, P., Defais, M., Samson, L. & Schendel, P. *Molec. gen. Genet.* **157**, 1–9 (1977); **162**, 299–305 (1978).
4. Schendel, P. F., Defais, M., Jeggo, P., Samson, L. & Cairns, J. *J. Bact.* **135**, 466–475 (1978).
5. Schendel, P. F. & Robins, P. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6017–6020 (1978).
6. Robins, P. & Cairns, J. *Nature* **279**, 74–76 (1979).
7. Riazuddin, S. & Lindahl, T. *Biochemistry* **17**, 2110–2118 (1978).
8. Singer, B. *Nature* **264**, 333–339 (1976).
9. Kirtikar, D. M. & Goldthwait, D. A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2022–2026 (1974).
10. Pegg, A. E. *Biochem. biophys. Res. Commun.* **84**, 166–173 (1978).
11. Jeggo, P. *J. Bact.* (submitted).

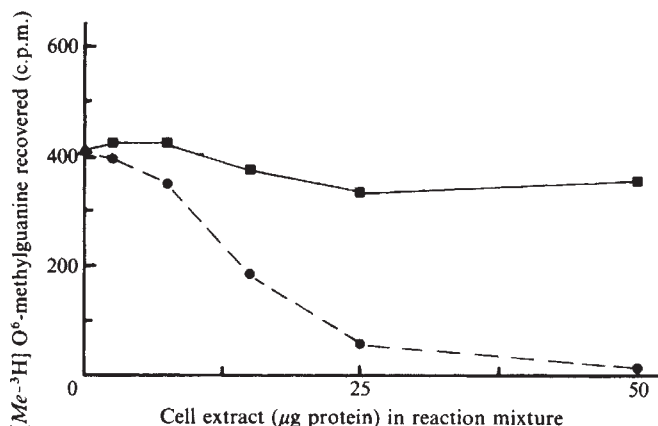


Fig. 2 Concentration dependence of the activity of an adapted *E. coli* cell extract on DNA containing O^6 -methylguanine residues. Experimental conditions and symbols are as in Fig. 1. The extracts were adjusted to $25\text{ mg protein per ml}$ and diluted in cold reaction mixture (without DNA) immediately before use; $2\text{ }\mu\text{l}$ of each dilution was added to the standard reaction mixture.

Hormonally specific phosphorylation of cardiac troponin I and activation of glycogen phosphorylase

WE recently reported that PGE₁ and isoprenaline have very different effects on the substrates of cyclic AMP-dependent protein kinase that regulate cardiac glycogen metabolism. Both agents elevated the cyclic AMP content and the protein kinase activity ratio in isolated perfused rat hearts, but only isoprenaline caused changes in the activity of the substrates of protein kinase (presumably phosphorylation, causing activation of phosphorylase kinase and inactivation of glycogen synthase)^{1,2}. Thus, the activation of soluble protein kinase appears insufficient to cause alteration of the regulatory enzymes of glycogen metabolism. This suggests that the glycogenolytic responses to isoprenaline involve actions other than the activation of soluble protein kinase. We now report the effects of PGE₁ and isoprenaline on another substrate of protein kinase, the inhibitory subunit of cardiac troponin (TNI). Since the phosphorylation of TNI may mediate the positive inotropic effects associated with β -adrenergic stimulation and cyclic AMP accumulation³, we felt that the hormonally specific expression of protein kinase activity might be used as a tool to determine whether similar specificity applies to the phosphorylation of TNI and to establish the functional relevance of that phosphorylation.

Although no biochemical consequence of TNI phosphorylation has been found to explain the positive inotropic action of drugs, some correlative data suggest a role for phosphorylation of TNI in the regulation of myocardial contractility. Cardiac TNI is phosphorylated in perfused hearts following exposure to β -adrenergic agonists and the phosphorylation correlates well with the time course and concentration-dependence of enhanced contractility³. When purified cardiac troponin is incubated with cyclic AMP-dependent protein kinase and [γ -³²P]ATP, the rate of phosphorylation is stimulated by cyclic AMP and inhibited by the heat stable inhibitor of protein kinase. The ³²P is incorporated specifically into the TNI subunit at rates two- to threefold greater than into mixed histones⁴. These results suggest that the rapid and specific phosphorylation of the I subunit of troponin may be a consequence of cyclic AMP accumulation in cardiac muscle. However, it is not clear whether phosphorylation of TNI always follows from elevation of intracellular cyclic AMP or results only from β -adrenergic stimulation^{3,5}. Accordingly, we have compared the effects of isoprenaline and PGE₁ on cyclic AMP content, protein kinase activity, the phosphorylation of TNI, and contractility in the isolated perfused rat heart.

Measurement of the phosphate content of TNI depends for its accuracy on the purity of the material isolated from the perfused hearts. A densitometer tracing of a gel electrophorogram of the material isolated by affinity chromatography over troponin C columns (Fig. 1) shows that the major band of protein (>90%) moved with an apparent molecular weight of 26,100, in good agreement with the published value for TNI from other laboratories^{3,4}. The purification by affinity chromatography and homogeneity with appropriate mobility on SDS-polyacrylamide gel electrophoresis are good criteria supporting the identity of the isolated protein as TNI. Small amounts of impurities (<10%) occasionally noticed on electrophorograms showed mobilities corresponding to those of TNC and TNT troponin subunits.

Both isoprenaline and PGE₁ increased the cyclic AMP content and the activity ratio of protein kinase in perfused rat hearts (Table 1). The activation of protein kinase occurred without changes in total activity (that measured in the presence of 2 μ M cyclic AMP). Only isoprenaline treatment increased the phosphate content of TNI, and caused activation of phosphorylase and an enhanced rate of ventricular pressure

development. PGE₁ did not show these effects despite its capacity to stimulate accumulation of cyclic AMP. The effects of 10 nM isoprenaline and 30 μ M PGE₁ on cyclic AMP accumulation and protein kinase activation were additive, whereas the activity ratio of phosphorylase and the degree of phosphorylation of TNI were the same following treatment with isoprenaline alone and with the combination of isoprenaline and PGE₁.

We have found, like England³, that treatment of perfused rat hearts with isoprenaline increases the phosphate content of TNI. Our values were up to 0.8 mol phosphate per mol TNI, in reasonable accord with the values reported by England for perfused rat heart³ and comparable to values for perfused mouse heart⁶. These measurements are lower than the phosphate content of TNI prepared from rabbit heart (1–1.9 mol phosphate per mol TNI) as reported by Perry and co-workers^{7,8} but are in good agreement with the *in vitro* experiments of Stull and Buss⁴ who, using purified components, found that cyclic AMP-dependent protein kinase catalysed the maximal incorporation of 1 mol phosphate per mol TNI, suggesting a single site available for phosphorylation per TNI molecule. However, we emphasise the relative changes in TNI phosphate and the failure of PGE₁ to cause TNI phosphorylation.

Previous reports have demonstrated that not all changes in cardiac contractility correspond to an altered phosphorylation state of TNI. For instance, England³ found that the inotropic effect of glucagon was maximal before significant TNI phosphorylation could be detected. Ezrailson *et al.*⁹ reported that in perfused cat heart ouabain, increased frequency of stimulation, and elevated external calcium induced positive inotropic responses with no changes in TNI-phosphate. Frearson *et al.*¹⁰ found no change in TNI phosphorylation when the inotropic state of perfused rabbit heart was altered by low-sodium and

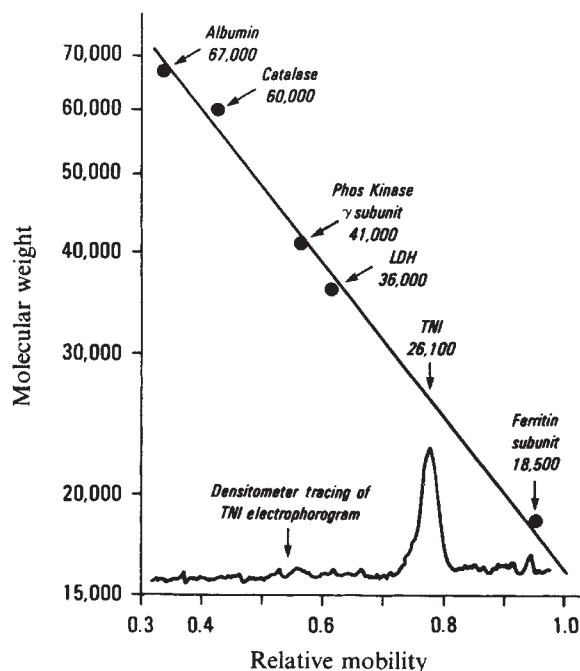


Fig. 1 SDS-polyacrylamide gel electrophoresis of TNI and calibrating proteins. Samples were solubilised as described by Stull and Buss⁴. Samples and standards (Pharmacia) were applied to 0.75 $\times$ 100 mm gel consisting of a 3.3% stacking gel and a 7.0% separating gel containing 0.5% SDS with Tris-glycine-SDS reservoir buffer. Following electrophoresis, gels were stained with Coomassie blue R (0.15%) in aqueous methanol (50%) acetic acid (7%), then destained with methanol acetic acid and equilibrated with acetic acid (7%). Stained bands were located and quantified by integration of areas under peaks of densitometer scans. The main figure shows a calibration curve (log of molecular weight versus mobility) indicating an apparent molecular weight of 26,100 for rat cardiac TNI. The inset shows actual densitometer tracing of gel with 2 μ g of rat TNI applied. Purity of >95% was observed when up to 100 μ g of TNI were applied.

Table 1 Effects of isoprenaline and PGE₁ singly and in combination

Treatment (n)	Cyclic AMP (pmol per mg protein)	Protein kinase (–cAMP: +cAMP)	Phosphorylase (–AMP: +AMP)	Phosphate content of TNI (mol per mol)	dP/dt
Control (7)	5.0 ± 0.2	0.17 ± 0.01	0.05 ± 0.02	0.34 ± 0.03	1.00
INE (10 nM) (7)	8.6 ± 0.5**	0.26 ± 0.01**	1.36 ± 0.02**	0.49 ± 0.04**	1.83**
INE (100 nM) (5)	—	0.39 ± 0.01**	0.58 ± 0.03**	—	2.20**
PGE ₁ (30 µM) (6)	7.3 ± 0.3**	0.24 ± 0.02**	0.08 ± 0.03	0.38 ± 0.06	0.94
INE (10 nM) + PGE ₁ (30 µM) (6)	11.2 ± 0.7**	0.33 ± 0.01**	0.34 ± 0.03**	0.53 ± 0.07*	1.77**
INE (100 nM) + PGE ₁ (30 µM) (4)	13.1 ± 1.1**	0.41 ± 0.02**	0.58 ± 0.03**	0.51 ± 0.05**	2.10**

Values are mean ± s.e.m. Asterisks indicate values differing significantly from control values by unpaired *t*-test analysis (*, $P < 0.02$; **, $P < 0.005$). dP/dt is expressed as a fraction of mean control value (1,680 ± 60 mm Hg s⁻¹). Hearts were excised from heparinised (1,500 U per kg), anaesthetised (pentobarbital, 100 mg per kg) male Sprague-Dawley rats (200–300 g) and immersed in ice cold saline until beating ceased. Hearts were perfused at 32 °C via the aorta without recirculation at an aortic pressure of 60 mm Hg with Krebs-Henseleit buffer as described by Neeley¹⁷. Mean coronary flow was 8 ml min⁻¹. Hearts were paced at 240 beats per min and equilibrated for 30 min before treatment. Left ventricular pressure and its rate of development (dP/dt) were monitored using an intraventricular needle connected to a pressure transducer. Following a 2-min perfusion with buffer containing drug or vehicle, hearts were frozen with Wollenberger clamps cooled in liquid N₂. The atria and great vessels were trimmed away and the remainder pulverised in a percussion mortar cooled in liquid N₂. The powder was stored at –70 °C until assayed. Cyclic AMP purified from acidic extracts of frozen tissue powder was quantified by the Gilman protein binding assay¹⁸. Protein kinase activity was assayed in the 300,000 g supernatant of a 1:10 homogenate of frozen tissue powder by the method of Keely *et al.*¹⁹. The activation state of protein kinase is expressed as the ratio of activities measured in the absence and presence of 2 µM cyclic AMP. Glycogen phosphorylase activity was assayed fluorometrically²⁰; activities are expressed as the ratio of activities assayed in the absence and presence of 2 mM AMP. We have previously described these assays in more detail². TNI was purified by troponin C affinity chromatography as described by Syska *et al.*²¹ on TNC columns. For this procedure, 500 mg of frozen heart powder were homogenised in the presence of 8M urea which was removed only after the TNI was purified, so that enzymatic modification during purification was minimised³. The yield was ~120 µg TNI per g heart and was independent of the extent of phosphorylation. Purity of TNI was ≥90% as established for each sample by SDS polyacrylamide electrophoresis as demonstrated in Fig. 1. Protein-bound phosphate was measured by a combination of the general assay for phosphoester bonds in proteins²² with the assay of inorganic phosphate of Itaya and Uj²³ and the ashing procedure of Ames²⁴, as detailed by Stull and Buss⁴. Protein was quantified by the method of Bradford²⁵ using bovine serum as a standard. We used a value of 23,550, as determined by amino acid analysis²⁶, as the molecular weight of TNI for our calculations.

high-calcium media. Nonetheless, the hypothesis that cardiac troponin may be phosphorylated *in vivo* in response to stimuli which increase cyclic AMP⁵ and that this phosphorylation may be associated with increased contractile force³ remains attractive. Our data suggest that this hypothesis must be modified since PGE₁ increases cyclic AMP but causes no net phosphorylation of TNI.

We have previously² discussed several possible explanations for this anomalous effect of PGE₁. A direct inhibitory effect of PGE₁ at a stage distal to protein kinase activation is unlikely since a high concentration of PGE₁ (30 µM) caused no inhibition of phosphorylase *a* formation, TNI phosphorylation, or the inotropic effects of 10 nM isoprenaline. We have not been able to detect an increase in phosphoprotein phosphatase activity due to PGE₁ (ref. 2). Maximally effective concentrations of the two agents produced the same activation of protein kinase as did isoprenaline alone, suggesting that we are not dealing with two populations of responsive cells. We must also consider the possibility that PGE₁ caused a transient change in protein phosphorylation such that our measurements, made on hearts freeze-clamped after 2 min of drug treatment, would not reflect the transient effect. Constant monitoring showed no transient change in dP/dt during the 2 min of treatment. The data of Keely (ref. 11 and personal communication) demonstrate that continuous perfusion with catecholamine or PGE₁ causes quick increases in cyclic AMP content and protein kinase activity ratios that are maintained for the duration of drug treatment (up to 10 min). During a 10-min treatment, catecholamine causes a sustained activation of phosphorylase whereas PGE₁ elicits no activation¹¹. Thus, it seems that PGE₁ causes no transient increase in protein phosphorylation. England has demonstrated the persistence of TNI-bound phosphate following a pulse of isoprenaline³. His result suggests that a transient activation of TNI phosphorylation might still be noticeable as increased TNI-bound phosphate at a later time. Consideration of these points argues against a transient change in protein phosphorylation in response to PGE₁.

Two alternative explanations are attractive: that β-adrenergic stimulation supplies a 'factor' that PGE₁ does not, or that PGE₁ and isoprenaline activate protein kinases in different cellular compartments. We have previously discussed these possibilities

with respect to the regulation of cardiac glycogen metabolism. If β-agonists supply a specific factor necessary for the expression of protein kinase activity, it would probably not be Ca²⁺, since protein kinase is not known to require this cation². In the present case, however, the discussion must be more tentative since the regulation of TNI phosphorylation and contractility by cyclic AMP-dependent protein kinase is by no means as clearly established as the involvement of protein kinase in the stimulation of glycogenolysis and because Ca²⁺ is intimately involved in the contractile process. The compartment of protein kinase within the cell and the activation of an ineffective pool of the kinase in response to PGE₁ are testable hypotheses. Several 'pools' of protein kinase are known—particulate and soluble, isozymes I and II (refs 12, 13); substrates and cyclic AMP, too, can be localised^{14,15}.

To test the possibility that the two isozymes of protein kinase might be selectively involved in the differing responses to isoprenaline and PGE₁, we have looked for hormonal specificity in hearts differing in their content of the isozymes—rat (>80% type I) and guinea pig (>90% type II)^{12,16}. In both of these preparations, isoprenaline and PGE₁ increase intracellular cyclic AMP and cause comparable elevations of the protein kinase activity ratio, but only isoprenaline causes the activation of glycogen phosphorylase and an enhanced dP/dt. Thus, activation of protein kinase and the hormonally specific expression of protein kinase activity occur in hearts in which either isozyme of protein kinase predominates (data not shown).

The present results confirm that TNI-bound phosphate increases in perfused rat heart treated with isoprenaline. The phosphorylation of TNI is functionally relevant to the extent that enhanced dP/dt and increased TNI-phosphate occur concurrently; and when, following exposure to PGE₁, we record no change in dP/dt, we find no change in the phosphate content of TNI.

Our study indicates that TNI phosphorylation and other sequelae are not the consequence merely of increased intracellular cyclic AMP. β-Adrenergic stimulation must provide a signal more complex than a generalised stimulation of cyclic AMP formation, since neither enhanced dP/dt nor increased phosphate content of TNI nor phosphorylase activation results from treatment of perfused rat heart with PGE₁, even though

PGE₁ increases both cardiac cyclic AMP content and the activity ratio of soluble protein kinase (Table 1). England's results³ suggest that glucagon may also provide the appropriate stimuli for enhanced protein phosphorylation and positive inotropism. This hormonally specific expression of protein kinase activity seems unrelated to the predominance of either of the isozymes of protein kinase. We are now extending these observations and considering the biochemical data in light of known functional and morphological compartments of the heart and in light of previous reports on functional pools of cyclic AMP and cyclic AMP-dependent protein kinases¹²⁻¹⁵. We are also aware that the compartment of cyclic AMP and protein kinases could have intriguing implications for the relationship of hormonal receptors, adenylate cyclase, and intracellular compartments.

We thank Michael Branks for technical assistance, Professor James T. Stull for loan of TNC affinity columns, Professor James W. Covell for his interest and Sandra Dutky for help in preparing the manuscript. Fellowships from the NIH (HL05363, to LLB) and the California Heart Association (to JSH) supported this work.

LAURENCE L. BRUNTON
J. SCOTT HAYES
STEVEN E. MAYER

Division of Pharmacology M-013,
University of California, San Diego,
La Jolla, California 92093

Received 19 March; accepted 3 May 1979.

1. Reese, J. B., Hayes, J. S., Brunton, L. L., Brown, J. H. & Mayer, S. E. *Pharmacologist* **20**, 232 (1978).
2. Hayes, J. S., Brunton, L. L., Brown, J. H., Reese, J. B. & Mayer, S. E. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1570-1574 (1979).
3. England, P. *Biochem. J.* **160**, 295-304 (1976).
4. Stull, J. T. & Buss, J. E. *J. biol. Chem.* **252**, 851-857 (1977).
5. Stull, J. T., Blumenthal, D. K., de Lanerolle, P., High, C. W. & Manning, D. R. *Adv. Pharmac. Therapeutics* **3**, 171-180 (1978).
6. England, P. J. *Biochem. J.* **168**, 307-310 (1977).
7. Cole, H. A. & Perry, S. V. *Biochem. J.* **149**, 525-533 (1975).
8. Moir, A. J. G. & Perry, S. V. *Biochem. J.* **167**, 333-343 (1977).
9. Ezrailson, E. G., Potter, J. D., Michael, L. & Schwartz, A. *J. molec. cell. Cardiol.* **9**, 693-698 (1977).
10. Frearson, N., Solaro, R. J. & Perry, S. V. *Nature* **264**, 801-802 (1976).
11. Keely, S. L. *Molec. Pharm.* **15**, 235-245 (1979).
12. Corbin, J. D., Sugden, P. H., Lincoln, T. M. & Keely, S. L. *J. biol. Chem.* **252**, 3854-3861 (1977).
13. Russell, D. H. *Adv. Cyclic Nucleotide Res.* **9**, 493-506 (1978).
14. Terasaki, W. L. & Brooker, G. *J. biol. Chem.* **252**, 1041-1050 (1977).
15. Parker, C. W. *Adv. Cyclic Nucleotide Res.* **9**, 647-660 (1978).
16. Hayes, J. S., Brunton, L. L. & Mayer, S. E. (in preparation).
17. Neely, J. R. & Rovetto, M. J. *Meth. Enzym.* **39**, 43-59 (1975).
18. Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **67**, 305-312 (1970).
19. Keely, S. L., Corbin, J. D. & Park, C. R. *J. biol. Chem.* **250**, 4832-4840 (1975).
20. Hardman, J. G., Mayer, S. E. & Clark, B. J. *Pharmac. exp. Ther.* **150**, 341-348 (1965).
21. Syska, M., Perry, S. V. & Trayer, I. P. *FEBS Lett.* **40**, 253-257 (1974).
22. Kleinsmith, L. J., Allfrey, V. G. & Mirsky, A. E. *Proc. natn. Acad. Sci. U.S.A.* **55**, 1182-1189 (1966).
23. Itaya, K. & Ui, M. *Clin. chim. Acta* **14**, 361-366 (1966).
24. Ames, B. N. *Meth. Enzym.* **8**, 116-118 (1966).
25. Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).
26. Grand, R. J. A., Wilkinson, J. M., Mole, L. E. *Biochem. J.* **159**, 633-641 (1976).

High-frequency generalised transduction by bacteriophage T4

GENERALISED transduction—the transfer of bacterial genes from one cell to another via bacteriophages—is one of a small number of processes by which bacteria can acquire exogenous genetic information. Transduction was first demonstrated in *Salmonella typhimurium* with the phage P22 by Zinder and Lederberg^{1,2} and subsequently in *Escherichia coli* with the phages P1³ and T1⁴. Bacteriophage T4 is the largest, and one of the most thoroughly studied of the virulent coliphages but has never been observed to mediate generalised transduction. In this report we describe a multiple mutant of T4 that displays the property of generalised transduction, and transfers *E. coli* genes with frequencies that, in general, are higher than those observed for other transducing phages.

The DNA of wild-type T4 contains glucosylated hydroxymethylcytosine (glu-HMC) in place of cytosine which enables

the phage to escape host-controlled DNA restriction. Multiply mutated derivatives of T4 have been isolated which are defective in several genes involved in the synthesis of HMC and the degradation of cytosine-containing DNA⁵. We examined one such mutant, constructed originally for molecular cloning purposes, to see if it had the ability to transduce bacterial DNA. Mutant T4GT7 (T4 generalised transducer no. 7) was isolated as a pseudorevertant of the triple mutant *amE51* (*g56*, *dCTPase*), *amC87* (*g42*; *dHMase*), NB5060 (deletion spanning *rIIB*, *denB*, *ac*) on a restriction deficient, non-amber suppressing (*Sup*^o) *E. coli* K host. The revertant carries an additional spontaneous *alc*⁻ mutation (E. Kutter, personal communication) and yields progeny with cytosine-containing DNA. The progeny DNA is sensitive to *in vivo* restriction by the plasmid-specified *EcoRI* system, and *in vitro* the DNA is cleaved by many restriction endonucleases⁶. When the mutant is grown on an amber-suppressing host (*Sup*⁺) permitting *g56* and *g42* to function, it produces progeny containing glu-HMC DNA which is resistant to restriction both *in vivo* and *in vitro*.

When auxotrophic strains of *E. coli* are infected with T4GT7 previously grown on bacteria carrying wild-type alleles, transductants arise at high frequency for most markers tested. Among the host markers for which we have observed transduction are the amino acids *leu* (2 min), *proA* (6 min), *his* (44 min), *arg G* (68 min) and *metB* (87 min), and the sugars *ara* (2 min), *lac* (8 min) and *xyl* (79 min) (ref. 7). In general, transductants arise at frequencies around 1 per 10⁴ singly infected cells, but marker-dependent variation is observed (Table 1). For some markers, notably *thr* (0 min), *malA* (74 min) and *mtl* (80 min) no transductants at all have been observed. Wild-type T4 is unable to transduce any of the markers we have tested; P1 does so, but at a much lower frequency than T4GT7, around 1 × 10⁻⁶. To date we have only observed co-transduction for the closely linked markers *ara* and *leu* which are co-transduced at the single marker frequency. The transducing capacity of T4GT7 lysates is not diminished by treatment with either DNase or RNase, eliminating the possibility that we are observing transformation rather than transduction. Finally, the number of transductants increases with increasing multiplicity of infection (MOI) through the range 0.001 to 0.5, then falls off sharply at higher MOIs as multiple infections reduce the number of singly infected, potentially transducible cells (Fig. 1). Together these data firmly establish that the mutant T4 phage are capable of generalised transduction.

In addition to transducing chromosomal markers, T4GT7 transduces extrachromosomal, plasmid-borne markers. Lysates

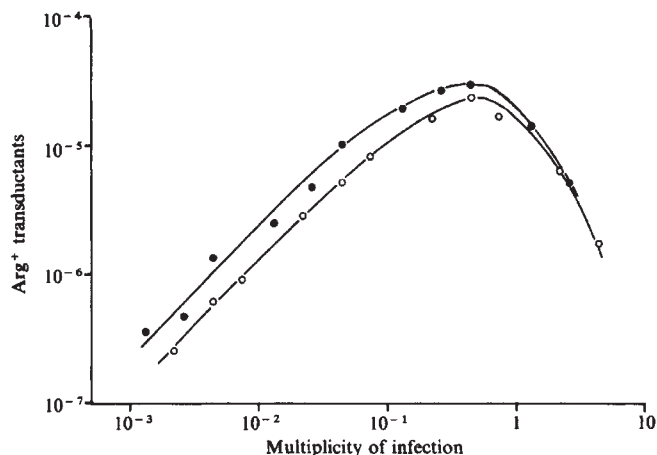


Fig. 1 Influence of multiplicity of infection on the formation of transductants. A culture of log-phase bacteria (JC411) was infected with T4GT7 at increasing multiplicities of infection and plated on minimal plates containing all the necessary supplements except arginine. The numbers of Arg⁺ transductants were counted and are plotted here as the ratio of transductants to total cells plated. Data from two different experiments using independently prepared phage lysates are shown.

Table 1 The frequencies with which T4GT7, T4 wild type and phage P1 transduce auxotrophic recipient strains JC 411 and AB 1157

JC 411				AB 1157			
<i>MetB</i>	<i>ArgG</i>	<i>His</i>	<i>Leu</i>	<i>Leu</i>	<i>Ara</i>	<i>ProA</i>	<i>Thr</i>
T4GT7 3.6×10^{-4}	7.1×10^{-5}	7.8×10^{-5}	1.1×10^{-4}	8.3×10^{-5}	5.9×10^{-5}	3.5×10^{-6}	$<10^{-7}$
P1 cm 1×10^{-6}	3×10^{-6}	3×10^{-6}	1×10^{-5}	7×10^{-6}	3×10^{-6}	10^{-7}	$<10^{-7}$
T4 wild type $<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$

The phage were grown on *E. coli* K ED8689 (*hsdR*⁻, *Sup*⁰) either as liquid or plate lysates. Recipient cells were grown to log phase in supplemented M9 minimal medium, resuspended in fresh medium in the presence of $20 \mu\text{g ml}^{-1}$ L-tryptophan to a titre of approximately 4×10^8 cells per ml and infected with sufficient phage to achieve a multiplicity of infection in the range 0.05–0.25. After 15 min adsorption at room temperature, aliquots of the infection mixtures were spread onto minimal selection plates and incubated for 2 days at 37 °C to allow growth of transductant colonies. Immediately before infection, the recipient cells were titred to determine the actual MOI. The data are shown as ratios of the number of transductants observed to the number of singly infected cells plated, the latter calculated from the Poisson formula knowing the cell titre and MOI. The absolute frequency with which markers are transduced varies two- to threefold from lysate to lysate. This fluctuation seems to be due to differences in the intracellular growth times of different phage preparations (Fig. 2). Lysates prepared on *Sup*⁺ donor bacteria have two- to threefold lower transducing capacities than lysates prepared on *Sup*⁰ hosts.

of the phage prepared on donors containing the antibiotic resistance-determining plasmids pBR322 or RP4 were found to transduce the recipient strain JC411 to ampicillin resistance. To standardise for donor-dependent differences in transducing capacity in these experiments, transduction for the chromosomal *arg* marker was simultaneously measured. Routinely, ampicillin resistance was found to be transduced twice as frequently as *Arg*⁺ from the pBR322-containing host, but only 1/20th as frequently from the RP4-containing host. As plasmid pBR322 exists in approximately 50 copies per cell⁸ whereas plasmid RP4 exists in only a few copies⁹, a difference in relative transduction frequency would be expected and is observed. The majority (170/171) of ampicillin-resistant transductants tested from the pBR322-grown lysate also were found to be tetracycline resistant, suggesting that the entire plasmid may be transduced although this has not been proved conclusively.

To determine at which stage during the development of T4GT7 transducing particles form, samples of a phage-infected bacterial culture were artificially lysed at various times after infection. The *Arg*⁺ transducing capacity reached a maximum 30 min after infection and declined thereafter (Fig. 2). The maximum for *Arg*⁺ transductants is about 10 times higher than the frequency observed in lysates that have been incubated for several hours. Most, but not all, other markers also showed an early maximum. The transducing capacity curve presumably reflects variation in the rates at which chromosomal fragments

and phage DNA molecules are packaged and accumulate and accounts for variation in transduction from lysate to lysate. Although the maximum transduction frequency is observed for the 30-min lysate, the formation of transducing particles does not cease at that time. Multiplying the transduction frequency by the viable phage titre at each time point during the lytic cycle indicates that transducing particles accumulate for at least an hour after infection and thereafter their number remains unchanged.

Possession of a circularly permuted phage chromosome and the ability to carry out generalised transduction often go hand in hand. Bacteriophage T4 is the classical circularly permuted phage¹⁰, and although the wild-type phage is unable to transduce, the multiply mutated derivative T4GT7, described here, can. We are not yet certain whether all the mutations present in the mutant are critical for transduction. Since suppression of the *amber* mutations in genes 42 and 56 reduces the transducing capacity only slightly, it is probable that these two mutations are irrelevant. The remaining mutations, *denB*⁻ and *alc*⁻, are almost certainly required for transduction since both affect host and phage DNA after infection. The *denB*⁻ mutation is a large deletion; in addition to abolishing the T4 cytosine-DNA-specific endonuclease IV, it also eliminates the *rIIIB* and *nad* functions¹¹, *nad* being involved in the disruption of the host nucleoid¹². The *alc* mutation affects transcription from cytosine containing templates⁵ and could also affect the *unf* function¹³, responsible for unfolding the bacterial chromosome, as other *alc*⁻ mutations are known to do¹⁴. Since T4GT7 was isolated as a plaque-forming derivative of the triple mutant, the possibility exists that it carries additional mutations beside *alc*⁻. Further analysis is under way to determine the precise genetics of T4 transduction.

At present, we do not know the form in which host DNA is transduced by T4GT7. Like P1¹⁵ and P22¹⁶, T4GT7 might transduce continuous sections of the host chromosome in linear, headful-sized pieces. Alternatively it might scramble the chromosome and package otherwise unlinked fragments in the same head; or, like phage Mu¹⁷, it may couple fragments of host and viral DNA and encapsulate hybrid molecules.

The observations presented here indicate that T4 is capable of efficient generalised transduction. Both the phage and its host are well understood from the genetic and physiological points of view^{7,18}. T4 is twice the size of P1 and might be expected to co-transduce correspondingly larger sections of the host chromosome, perhaps up to 5-min as compared with 2-min for P1. The increased capacity, together with its higher transducing frequency, recommend T4GT7 as an alternative to P1 for practical laboratory use and other applications.

GEOFFREY G. WILSON
KAREN K. Y. YOUNG
GORDON J. EDLIN

Department of Genetics, University of California,
Davis, California 95616

WILLIAM KONIGSBERG
Department of Molecular Biophysics and Biochemistry,
Yale University, New Haven, Connecticut

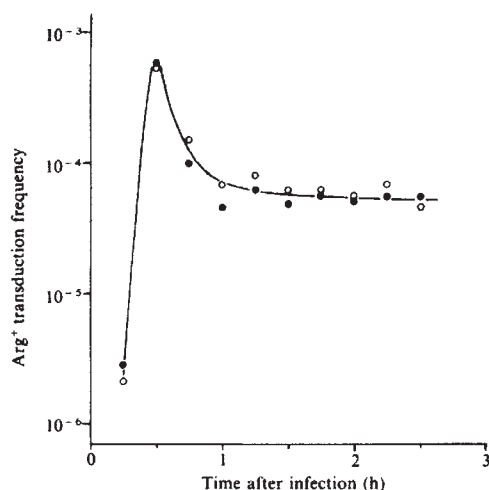


Fig. 2 Kinetics of transducing particle formation. A log-phase culture of ED8689 was infected at time zero with T4GT7 at a MOI of 5. Ten minutes later the culture was superinfected with the same multiplicity of phage to induce lysis inhibition. Starting at 15 min, and at 15-min intervals thereafter, samples of the infected cells were taken, lysed with chloroform, and the plaque-forming progeny titred. The infected cells were incubated at 37 °C with aeration throughout the experiment. Once the titre had been established, the *Arg*⁺ transducing capacity of each sample was determined with JC 411 at MOI 0.5. Two independent estimates were made and both sets of data are shown.

Received 7 March; accepted 11 May 1979.

1. Zinder, N. D. & Lederberg, J. *J. Bact.* **64**, 679–699 (1952).
2. Susskind, M. M. & Botstein, D. *Bact. Rev.* **42**, 385–413 (1978).
3. Lennox, E. S. *Virology* **1**, 190–206 (1955).
4. Drexler, H. *Proc. natn. Acad. Sci. U.S.A.* **66**, 1083–1088 (1970).
5. Snyder, L., Gold, L. & Kutter, E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3098–3102 (1976).
6. Wilson, G. G. *et al. Molec. gen. Genet.* **156**, 203–214 (1977).
7. Bachmann, B. J. *et al. Bact. Rev.* **40**, 116–167 (1976).
8. Rodriguez, R. L. *et al. Miami Winter Symp.* **13**, 73–84 (1977).
9. Shapiro, J. A. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A., Shapiro, J. & Adhya, S.) 630 (Cold Spring Harbor Laboratories, New York, 1977).
10. Streisinger, et al. *Proc. natn. Acad. Sci. U.S.A.* **51**, 775–779 (1964).
11. Depew, R. E. *et al. Virology* **64**, 144–152 (1975).
12. Snustad, D. P. & Conroy, L. M. *J. molec. Biol.* **89**, 663–673 (1974).
13. Snustad, D. P. *et al. J. Virol.* **17**, 622–641 (1976).
14. Sirotkin, K. *et al. Nature* **265**, 28–32 (1977).
15. Ikeda, H. & Tomizawa, J. I. *J. molec. Biol.* **14**, 85–109 (1965).
16. Ebel-Tsipis, J. *et al. J. molec. Biol.* **71**, 433–448 (1972).
17. Chow, L. T. & Bukhari, A. I. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A., Shapiro, J. & Adhya, S.) 295–306 (Cold Spring Harbor Laboratory, New York, 1977).
18. Wood, W. B. & Revel, H. R. *Bact. Rev.* **40**, 847–868 (1976).

Inability of circular mRNA to attach to eukaryotic ribosomes

IN contrast to prokaryotic systems, initiation by eukaryotic ribosomes seems to be restricted to a single site near the 5'-terminus of the message. However, nucleotide sequence analyses have not revealed a common 'recognition sequence' within the 5'-proximal portion of eukaryotic messages. We have previously suggested that the position of an AUG codon within a eukaryotic message—rather than the surrounding nucleotide sequence—might be the primary feature defining an initiation site. We proposed^{1,2} that a 40S ribosomal subunit binds initially at or near the 5'-end of a message and migrates along the RNA chain until the first AUG codon is encountered, whereupon it stops, a 60S subunit joins, and polypeptide synthesis begins. Although this mechanism accounts for many of the peculiarities of eukaryotic translational initiation, it leaves unanswered the question of how ribosomes recognise the 5'-terminus of mRNA. The m⁷G cap which blocks the end of most eukaryotic messages³ is not responsible for directing ribosomes to the 5'-terminus since ribosomes initiate at the 5'-proximal AUG even in uncapped or unmethylated RNAs^{4,5}. The ability of ribosomes to select the 5'-end of a message irrespective of its chemical structure (that is, whether the terminus is m⁷GpppN..., GpppN... or pppN...) is difficult to explain. An intriguing possibility is that eukaryotic ribosomes attach to mRNA by a threading mechanism. A simple prediction of the threading mechanism is that circularisation of a message should abolish its ability to interact with ribosomes. The experiments described below fulfill this prediction. The possibility of threading by prokaryotic ribosomes was ruled out, several years ago, by the demonstration that *Escherichia coli* ribosomes were able to translate the circular DNA of phage fd^{6,7}. Indeed, the simplest version of a threading mechanism (entry at the 5'-terminus of a message with translation beginning at the first AUG codon encountered by the ribosome) is contradicted by the ability of *E. coli* ribosomes to initiate at multiple internal sites in polycistronic messages, such as bacteriophage R17 RNA. Conversely, the threading mechanism is consistent with the monocistronic character of eukaryotic messages.

These experiments were designed to determine whether mRNA molecules retain the capacity to attach to eukaryotic ribosomes after circularisation of the message, catalysed by T4 RNA ligase⁸. Templates were sought which were able to bind with reasonable efficiency to both prokaryotic and eukaryotic ribosomes, and were small enough to undergo circularisation. Two kinds of synthetic polymers met these requirements. One series of experiments was carried out with linear and circular molecules containing a single AUG codon flanked by polyadenylic acid. A random copolymer of adenylic, uridylic and guanylic acid was used for another set of experiments.

Circular ³²P-labelled molecules containing a single AUG codon were synthesised as follows. (1) The dinucleotide UpG

was converted to UpG(pA)_n in a reaction catalysed by primer-dependent polynucleotide phosphorylase⁹ (PNPase), with ³H-ADP as precursor. In the presence of 0.5 M NaCl, a homo-

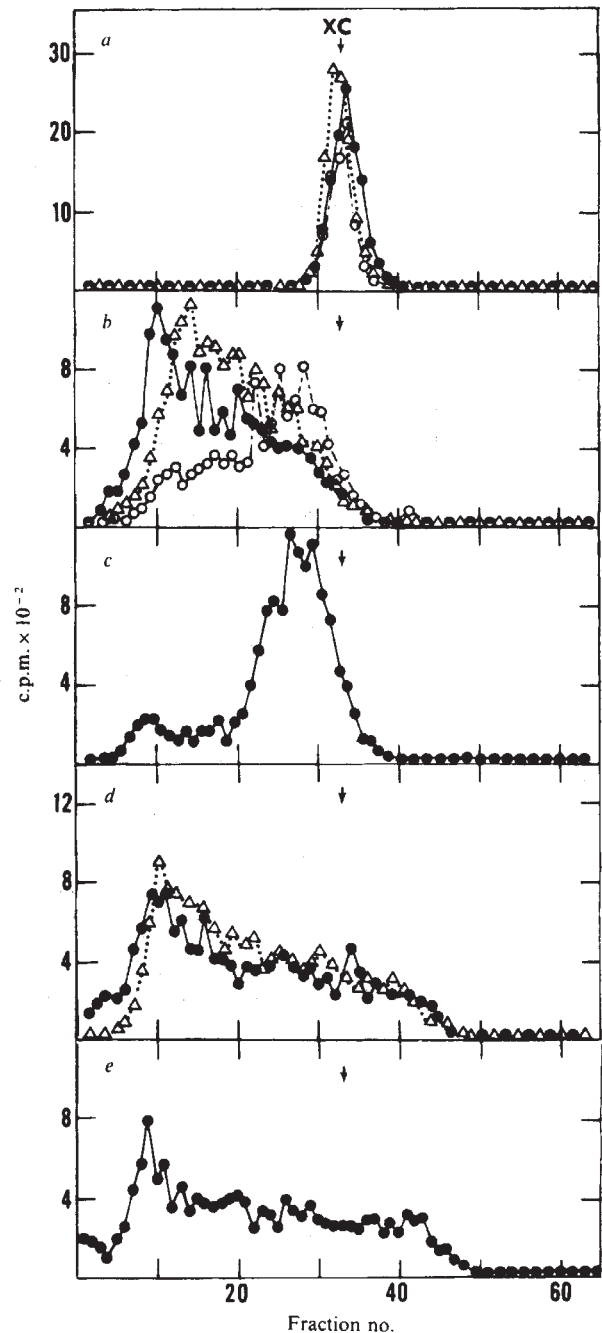


Fig. 1 Polyacrylamide gel electrophoresis of ³²P-labelled ribopolymers. The origin is at the left. Three forms of the polymer UpG(pA)₂₅ were analysed in separate gels, and the three profiles are superimposed in *a*: linear [⁵-³²P]pUpG(pA)₂₅ before circularisation (Δ...Δ); after reaction with RNA ligase (●—●); and after cleavage of the circular molecules with T₁ RNase (○—○). Incubation with T₁ RNase (40 U enzyme and 20 μg of RNA per 60 μl reaction) was for 20 min at 37 °C, followed by phenol-SDS extraction and ethanol precipitation. The same symbols in *b* represent three forms of the polymer UpG(pA)₁₁₀. After reaction with RNA ligase, an aliquot of UpG(pA)₁₁₀ was further treated with PNPase to degrade residual linear molecules (*c*). Electrophoresis of [⁵-³²P]poly(A, U, G) linear (Δ...Δ) and circular (●—●) molecules is shown in *d*. An aliquot of purified poly(A, U, G) circles, from which linear molecules had been removed with PNPase, was analysed in *e*. Composition of the gels and electrophoresis conditions were as described previously¹⁰. Arrows show the position of xylene cyanol FF blue (XC). Other markers were bromophenol blue (fraction 61) and transfer RNA (fraction 21).

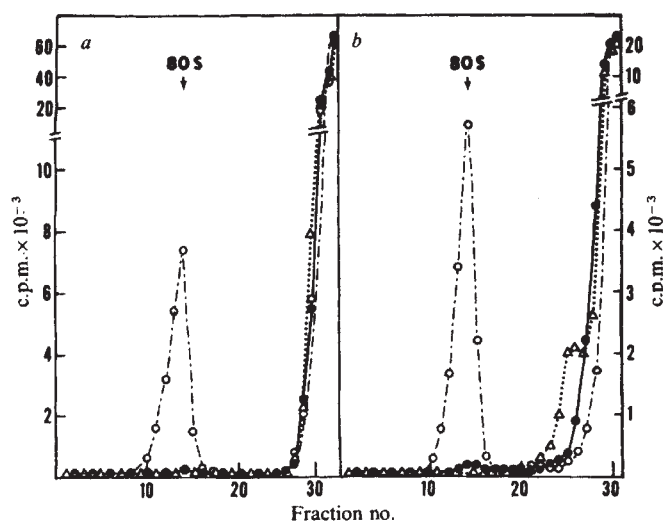


Fig. 2 Binding of ^{32}P -labelled pUpG(pA) $_n$ polymers to wheat-germ ribosomes. *a*, Binding of three forms of the polymer pUpG(pA) $_{25}$ (150,000 c.p.m. per 50 μl reaction); *b*, binding of three forms of the polymer pUpG(pA) $_{110}$ (80,000 c.p.m. per 50 μl reaction). Polymers were tested for ribosome binding before reaction with RNA ligase ($\Delta \dots \Delta$), after circularisation with RNA ligase ($\bullet \dots \bullet$), and after cleavage of the circles with T_1 RNase ($\circ \dots \circ$). Reaction conditions with wheat-germ S23 extract were as described previously^{2,10}. Following 6 min incubation at 19 $^\circ\text{C}$, samples were chilled and layered on to 10–30% glycerol gradients. Centrifugation was for 3 h in a SW41 rotor. The direction of sedimentation was from right to left in all gradients. Gradient fractions (0.37 ml) were mixed with Beckman GP scintillation cocktail and counted for radioactivity. In each panel, profiles obtained from three parallel gradients have been superimposed.

geneous polymer containing ~ 25 adenylic acid residues was obtained (Fig. 1*a*, triangles). Higher molecular weight polymers (which were also more heterogeneous in size, Fig. 1*b*, triangles) were obtained by lowering the NaCl concentration¹¹. (See Table 1 for reaction conditions.) Average chain lengths of 25 and 110 nucleotides for UpG(pA) $_n$ preparations 1 and 2, respectively, were determined from the ratio of ^3H -labelled adenosine/3'-AMP, following alkaline hydrolysis (not shown). (2) The linear molecules of UpG(pA) $_n$ were labelled at the 5'-terminus by phosphate transfer from [$\gamma\text{-}^{32}\text{P}$]ATP, catalysed by polynucleotide kinase. (3) Subsequent incubation with RNA ligase converted a substantial fraction of the 5'-terminal label to a form resistant to alkaline phosphatase (Table 1, 1*b* and 2*b*). As expected, the efficiency of ligation was greater with the shorter polymers (preparation 1) than with the longer molecules in preparation 2. Incorporation of 5'- ^{32}P into an internal phosphodiester bond, suggested by resistance to alkaline phosphatase, was confirmed by analysing the products of alkaline hydrolysis. Whereas the linear polymer [$5'\text{-}^{32}\text{P}$]pUpG(pA) $_n$ yielded pUp as the only ^{32}P -labelled product (Table 1), after incubation with RNA ligase ^{32}P -radioactivity from preparation 1 was quantitatively recovered as 3'-AMP (Table 1). Thus, preparation 1*b* seemed to consist entirely of circular molecules. Preparation 2*b*, on the other hand, clearly contained residual linear monomers (and perhaps linear dimers) as only 43% of the ^{32}P -radioactivity was recovered as 3'-AMP after alkaline hydrolysis. The residual linear molecules were eliminated by treatment with PNPase, yielding the purified circle preparation 2*c* in Table 1. (Comparison of the polyacrylamide gel profiles of preparations 2*a* (Fig. 1*b*, triangles) and 2*c* (Fig. 1*c*) reveals that the smaller molecules in the linear preparation were preferentially circularised. Although the average chain length of the linear molecules was 110 nucleotides, the majority of the circular molecules were in the range 40–80 nucleotides. Thus, preparations 2*b* and *c* are represented as pUpG(pA) $_{40-80}$ in the

following paragraphs.) (4) Finally, an aliquot of each pUpG(pA) $_n$ circle preparation was treated with ribonuclease T_1 , which cleaves on the 3'-side of guanosine residues, thereby converting each circle into a linear molecule terminating in AUG. This provided a useful control for the ribosome binding experiments described below and the treatment with T_1 RNase also confirmed that putative pUpG(pA) $_n$ circles had not been randomly cleaved by contaminating nuclease(s) at any stage (as the size of the linear (Ap) $_{25}$ UpGp molecules obtained after T_1

Table 1 Characterisation of linear and circular ribopolymers

^{32}P -labelled polymer preparation	Resistance to alkaline phosphatase	Resistance to polynucleotide phosphorylase	Products of alkaline hydrolysis	
			pUp	3'-AMP
1 <i>a</i> Linear [$5'\text{-}^{32}\text{P}$]pUpG(pA) $_{25}$	0.2%	not done	100%	0%
1 <i>b</i> [$5'\text{-}^{32}\text{P}$]pUpG(pA) $_{25}$ after reaction with RNA ligase	88%	90%	1%	99%
2 <i>a</i> Linear [$5'\text{-}^{32}\text{P}$]pUpG(pA) $_{110}$	0.3%	5%	100%	0%
2 <i>b</i> [$5'\text{-}^{32}\text{P}$]pUpG(pA) $_{110}$ after reaction with RNA ligase	29%	25%	57%	43%
2 <i>c</i> Circular ^{32}P -labelled pUpG(pA) $_{110}$ after hydrolysis of residual linear molecules with PNPase	80%	(100%)	11%	89%
3 <i>a</i> Linear [$5'\text{-}^{32}\text{P}$]poly(A, U, G)	0.4%	10%	not done	
3 <i>b</i> [$5'\text{-}^{32}\text{P}$]poly(A, U, G) after reaction with RNA ligase	76%	73%	not done	

Polymer preparation 1*a* was synthesised in a 200- μl reaction mixture containing 0.2 M Tris-Cl (pH 8.3), 0.5 M NaCl, 10 mM magnesium acetate, 0.7 mM UpG, 25 mM ^3H -ADP (1.8 mCi mmol $^{-1}$, NEN) and 1.25 U primer-dependent PNPase (PL Biochemicals). After incubation for 6 h at 34 $^\circ\text{C}$ the product was extracted with phenol, and unincorporated precursors were removed by filtration through Sephadex G-25. Similar conditions were used for preparation 2*a*, except that the concentration of UpG primer was 0.2 mM, and 0.2 M NaCl was used. Linear polymers (1*a*, 2*a*, 3*a*) were labelled at the 5'-terminus in reaction mixtures (125 μl) containing 36 mM Tris-Cl (pH 8), 120 mM NaCl, 10 mM MgCl_2 , 10 mM dithiothreitol (DTT), 3.5 μg bovine serum albumin (BSA), 50 μg of polymer, 40 μCi of [$\gamma\text{-}^{32}\text{P}$]ATP (33 Ci mmol $^{-1}$, NEN) and 3 U polynucleotide kinase (PL Biochemicals). Following incubation for 60 min at 37 $^\circ\text{C}$ the labelled polymers were extracted with phenol and filtrated through Sephadex G-25. Circular polymers (1*b*, 2*b*, 3*b*) were obtained by reaction with T_4 RNA ligase, using the low temperature conditions suggested by Bruce and Uhlenbeck¹². Reaction mixtures (450 μl) contained 1.6×10^7 c.p.m. of 5'-end labelled polymer, 11 U of RNA ligase (PL Biochemicals), 14 μg BSA, 10% (v/v) dimethyl sulphoxide, 50 mM HEPES (pH 7.6), 10 mM MgCl_2 , 4 mM DTT and 0.2 mM ATP. Incubation was for 19–24 h at 4 $^\circ\text{C}$. Hydrolysis of linear molecules with PNPase (preparation 2*c* and column 3) was accomplished by incubation for 40 min at 35 $^\circ\text{C}$ with 15 U ml $^{-1}$ of primer-independent PNPase (type 15, PL Biochemicals) in 50 mM Tris-Cl (pH 8.8), 175 mM NaCl, 10 mM K_2HPO_4 , 10 mM MgCl_2 , 1 mM EDTA and 17 μg BSA per ml. Carrier RNA was present at 100 μg ml $^{-1}$. For preparation 2*c*, the reaction was terminated by extraction with phenol-SDS, and low molecular weight products were removed by filtration through Sephadex G25. For the analytical PNPase analyses in column 3, radioactive polymers were precipitated with 5% trichloroacetic acid (TCA) at 4 $^\circ\text{C}$ and then counted on Millipore filters. Resistance to alkaline phosphatase (column 2) is based on TCA-precipitable radioactivity following incubation at 37 $^\circ\text{C}$ for 40 min with bacterial alkaline phosphatase (2 U ml $^{-1}$, Worthington), in 20 mM Tris-Cl (pH 8). The products of alkaline hydrolysis were identified by high voltage paper electrophoresis following incubation in 0.5 M KOH at 37 $^\circ\text{C}$ for 18 h.

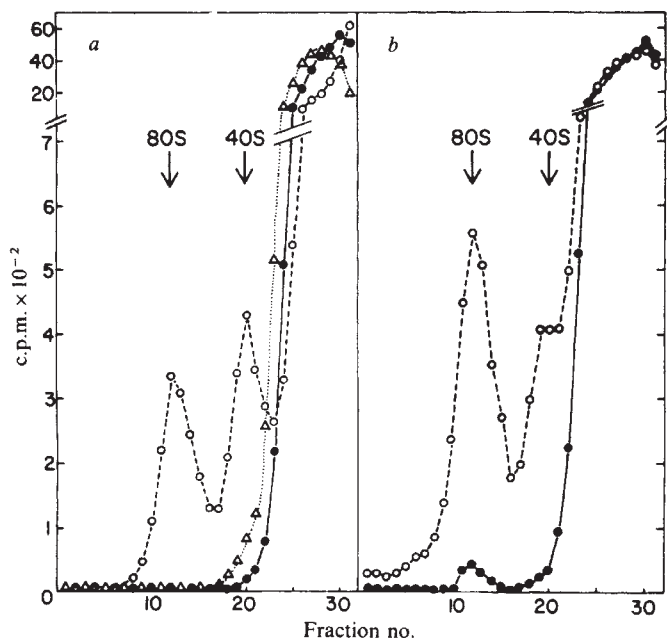


Fig. 3 Binding of linear and circular ^{32}P -labelled polymers to reticulocyte ribosomes. *a*, Three forms of the polymer pUpG(pA)_{110} were incubated with rabbit reticulocyte ribosomes: $\triangle$... $\triangle$, linear molecules before ligation; $\bullet$... $\bullet$, circles; $\circ$... $\circ$, circles which had been cleaved with T_1 RNase. Reaction conditions were as described previously¹³. Incubation was for 8 min at 32°C . Profiles from three parallel gradients have been superimposed. In *b*, linear ($\circ$... $\circ$) and circular ($\bullet$... $\bullet$) forms of poly (A, U, G) were tested. Each 200- μl reaction mixture contained 30,000 c.p.m.

RNase digestion of the circle preparation in Fig. 1*a* (open circles) was identical to that of the starting pUpG(pA)_{25} linear molecules (triangles). With the second preparation, the average size of the T_1 RNase-treated material (Fig. 1*b*, open circles) was smaller than the original linear pUpG(pA)_{110} molecules (Fig. 1*b*, triangles), consistent with the interpretation that only the smaller chains had been circularised. The third polymer preparation used in these studies was a random copolymer of adenylic, uridylic and guanylic acid. Poly(A, U, G) (Miles) was treated with alkaline phosphatase to generate a 3'-hydroxyl terminus and was subsequently labelled at the 5'-end with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Table 1, 3*a*). An aliquot of $[\gamma\text{-}^{32}\text{P}]\text{poly(A, U, G)}$ was incubated with RNA ligase, resulting in $\sim 75\%$ circularisation (Table 1, 3*b*). For some of the ribosome binding studies described below, residual linear molecules in the poly(A, U, G) circle preparation were hydrolysed with PNPase. The resulting purified poly(A, U, G) circle preparation was 90% resistant to alkaline phosphatase. Figure 1*d* and *e* show that, although there was considerable heterogeneity in the size of the poly(A, U, G) circles, most of the molecules were larger than 80 nucleotides.

Linear and circular forms of polymers containing a single AUG codon, flanked by poly(A), were tested for ability to interact with wheat-germ (Fig. 2), reticulocyte (Fig. 3*a*) and *E. coli* ribosomes (Fig. 4*a*). One reason for the selection of $[\gamma\text{-}^{32}\text{P}]\text{pUpG(pA)}_{25}$ as substrate for the RNA ligase reaction is that, before circularisation, the polymer does not contain an AUG codon. Thus, the linear precursor showed no detectable stable binding to wheat-germ ribosomes (Fig. 2*a*, triangles). Although circularisation of pUpG(pA)_{25} creates an AUG initiator codon, circular pUpG(pA)_{25} was also unable to bind to wheat-germ ribosomes (Fig. 2*a*, closed circles). When the circles were hydrolysed with RNase T_1 , however, the resulting linear molecules were able to form stable 80S complexes (Fig. 2*a*, open circles). Thus, inability of wheat-germ ribosomes to bind to circular pUpG(pA)_{25} cannot be attributed to absence of a proper initiation sequence. The sequence $(\text{Ap})_n\text{UpGp}$ (in a linear molecule) is adequate for stable attachment of wheat-

germ ribosomes, although the level of binding ($\sim 15\%$) was lower than would be observed with most natural messages. Failure of pUpG(pA)_{25} circles to interact with wheat-germ ribosomes might, however, be due to the small size of the circles. Indeed, the preparation of circles tested with wheat-germ ribosomes in Fig. 2*a* also failed to bind stably to *E. coli* ribosomes (not shown). Thus, the experiment in Fig. 2*a* was repeated using bigger polymers. The gradient profile in Fig. 2*b* (closed circles) shows only a very low level of binding ($< 1\%$) when circular molecules of the form pUpG(pA)_{40-80} were incubated with wheat-germ ribosomes. In contrast, 16% of the linear molecules resulting from T_1 RNase digestion (Fig. 2*b*, open circles) bound to wheat-germ 80S ribosomes. Figure 3*a* shows that reticulocyte ribosomes, like the wheat-germ system, discriminated against circular forms of the polymer pUpG(pA)_{40-80} .

In control experiments, linear and circular molecules of pUpG(pA)_{40-80} were tested for the ability to bind to *E. coli* ribosomes. Even without an AUG codon linear $[\gamma\text{-}^{32}\text{P}]\text{pUpG(pA)}_{110}$ showed a fairly high level of binding to *E. coli* ribosomes, varying from 3 to 10% of the input polymer. Therefore, to show that circular molecules can attach to *E. coli* ribosomes, binding had to be carried out using a purified preparation of pUpG(pA)_{40-80} circles (Table 1, line 2*c*) which was free of residual linear molecules. Figure 4*a* (closed circles) shows that 12% of the purified circles formed 70S initiation complexes with *E. coli* ribosomes which were fixed with glutaraldehyde¹⁴ before centrifugation. Binding to *E. coli* ribosomes was also detectable although at a reduced level without glutaraldehyde fixation (Fig. 4*a*, open circles). In contrast, the

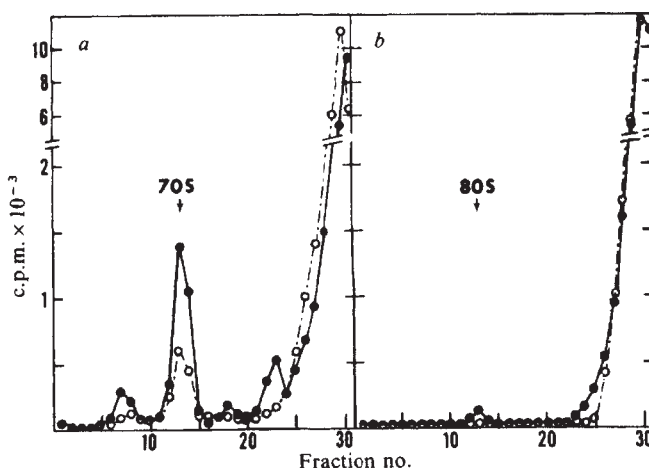


Fig. 4 Binding of purified ^{32}P -labelled pUpG(pA)_{40-80} circles to: *a*, *E. coli*; and *b*, wheat-germ ribosomes. Each 50- μl reaction mixture contained 60,000 c.p.m. of the polymer described in 2*c* of Table 1. For binding to *E. coli* ribosomes (*a*), the 50- μl reaction mixture contained 50 mM HEPES(NH_4), pH 7.5, 100 mM NH_4Cl , 6.6 mM magnesium acetate, 0.1 mM EDTA, 0.6 mM DTT, 0.5 mM GTP, 55 μg of *E. coli* MRE600 tRNA charged with methionine in formylating conditions, and 2.3 A_{260} units of low salt washed *E. coli* Q13 ribosomes. The ribosomes (retaining initiation factors, but free of elongation factors) were obtained by pelleting from an S30 extract and washing twice with 20 mM HEPES (pH 7.5), 60 mM NH_4Cl , 10 mM magnesium acetate and 6 mM β -mercaptoethanol. Incubation of the ^{32}P -labelled polymer with *E. coli* ribosomes was for 8 min at 35°C followed by sedimentation through a 10–30% sucrose gradient containing 20 mM HEPES (pH 7.6), 50 mM NH_4Cl , 7 mM magnesium acetate, 0.1 mM EDTA and 6 mM β -mercaptoethanol. Centrifugation was for 3.25 h at 4°C in the SW41 rotor. Reaction conditions used with wheat-germ ribosomes (*b*) were as described in Fig. 2. After incubation with either wheat-germ or *E. coli* ribosomes, half of each sample was directly applied to a gradient ($\circ$... $\circ$); the remainder was fixed by addition of 7 μg BSA and 2 vol 2.5% glutaraldehyde in 50 mM HEPES (pH 7.5), 50 mM KCl and 5 mM magnesium acetate. After 5 min at 4°C the glutaraldehyde-treated samples ($\bullet$... $\bullet$) were layered on to gradients and centrifuged.

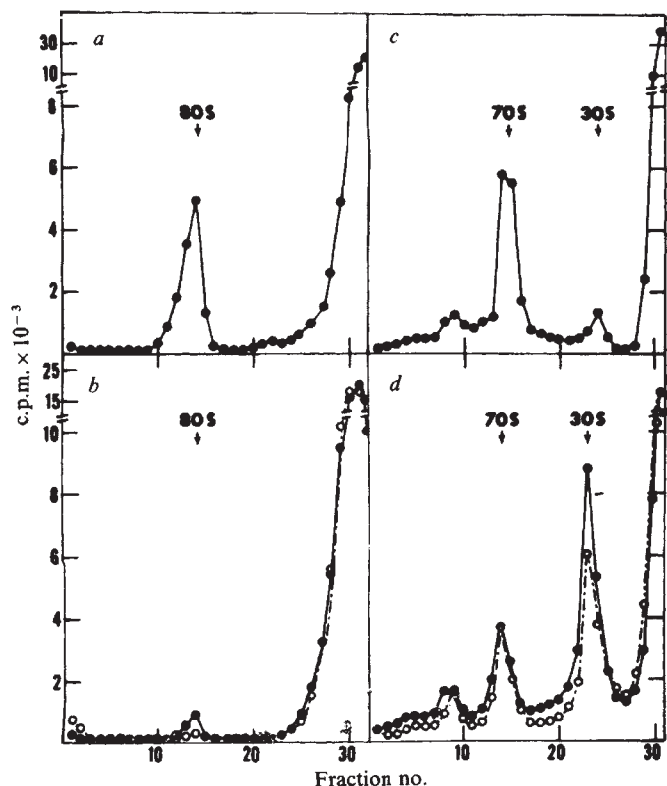


Fig. 5 Binding of ^{32}P -labelled linear and circular poly(A, U, G) to wheat-germ and *E. coli* ribosomes. Reaction conditions were as described in Figs 2 and 4, and initiation complex formation was assayed by centrifugation through gradients. Binding of linear $[5'\text{-}^{32}\text{P}]\text{poly(A, U, G)}$ is shown with wheat-germ ribosomes (a) and with *E. coli* ribosomes (c). Circular poly(A, U, G) was incubated with wheat-germ ribosomes in b and with *E. coli* ribosomes in d. In b and d, binding is shown for the poly(A, U, G) circle preparation described in line 3b of Table 1 (●—●) and also for a more highly purified preparation (○—○) from which residual linear molecules had been removed by hydrolysis with PNPase. Each reaction mixture contained 75,000 c.p.m. Samples were not fixed with glutaraldehyde.

same preparation of purified pUpG(pA)₄₀₋₈₀ circles showed no stable interaction with wheat-germ ribosomes, with or without glutaraldehyde fixation (Fig. 4b).

Another series of ribosome binding experiments was carried out with linear and circular forms of poly(A, U, G). Linear $[5'\text{-}^{32}\text{P}]\text{poly(A, U, G)}$ bound with approximately the same efficiency to wheat-germ (Fig. 5a) as to *E. coli* ribosomes (Fig. 5c). Circularisation of the template strikingly reduced its capacity to bind to wheat-germ ribosomes (Fig. 5b, closed circles). The very low (3%) level of binding to wheat-germ ribosomes was reduced even further when residual linear molecules were eliminated from the circle preparation (Fig. 5b, open circles). Even after fixation with glutaraldehyde, binding of circular poly(A, U, G) to wheat-germ ribosomes could not be detected (data not shown). Poly(A, U, G) circles also failed to form stable complexes with reticulocyte ribosomes (Fig. 3b, closed circles). In contrast, circular poly(A, U, G) attached to *E. coli* ribosomes (Fig. 5d) even more efficiently than the corresponding linear template. The reasons for the enhanced binding of circular templates, and for accumulation of some of the circular poly(A, U, G) molecules in 30S complexes, are not known.

The validity of the control experiments with *E. coli* ribosomes requires that the circular polymers are not nicked by endonucleases which might contaminate the *E. coli* translational system. Three observations suggest that circular RNA templates were not converted to a linear form before binding. (1) As mentioned previously, circular pUpG(pA)₂₅ failed to bind to *E.*

coli (as well as to wheat-germ) ribosomes, presumably due to the small size of the circles. The same polymer preparation was able to attach to *E. coli* ribosomes after cleavage of the circles with T₁ RNase. Thus, the level of endonuclease activity associated with *E. coli* ribosomes was apparently insufficient to convert circular pUpG(pA)₂₅ to a linear form. (2) The pattern of binding of poly(A, U, G) circles to *E. coli* ribosomes was qualitatively different from the binding of linear poly(A, U, G). That is, a large proportion of the circular molecules accumulated in 30S complexes. Although the reason for this is not known, the different binding patterns obtained with linear and circular poly(A, U, G) suggests that the circular molecules were not cleaved during the incubation. (3) Attachment of both linear and circular poly(A, U, G) templates to *E. coli* ribosomes was complete within 2 min of incubation. The rapid binding suggests that conversion of circular molecules to a linear form is not a prerequisite for attachment to *E. coli* ribosomes.

The present experiments suggest that eukaryotic ribosomes might thread on to the 5'-end of mRNA. Even in the absence of the m⁷G cap, the 5'-terminus of mRNA apparently has an essential role in ribosome attachment, as circularisation of pUpG(pA)₄₀₋₈₀ or of poly(A, U, G) totally abolished ability of the template to bind to wheat-germ or reticulocyte ribosomes. It seems unlikely that the size of the circles restricts their ability to interact with eukaryotic ribosomes. Although the small size of the pUpG(pA)₂₅ circle preparation may have precluded interaction with ribosomes, the larger pUpG(pA)₄₀₋₈₀ circles also failed to bind to wheat-germ or reticulocyte ribosomes. In the case of the circular poly(A, U, G) preparation used in Fig. 5, most of the molecules were more than 80 nucleotides long. Other trivial explanations for the inactivity of circular templates with eukaryotic ribosomes seem to be ruled out by two controls: (1) linear forms of each polymer were able to form initiation complexes in wheat-germ and reticulocyte extracts; and (2) the circular molecules that failed to bind to wheat-germ or reticulocyte ribosomes were able to bind to *E. coli* ribosomes. The striking difference in activity of circular templates with bacterial and eukaryotic ribosomes supports the suggestion¹ that the mechanism by which eukaryotic ribosomes select initiation regions in mRNA seems to be quite different from that of prokaryotes.

If mRNA threads through an aperture on the 40S ribosome, one might anticipate that the 40S subunit would be unable to detach when the ribosome arrives at a termination codon. Following release of the 60S subunit and the nascent polypeptide, the 40S subunit might remain on the message until it reaches the 3'-terminus. An alternative possibility is that the aperture through which the message threads is formed by the junction of the 40S subunit with initiation factor(s), which detach after mediating the early steps of translation.

I thank Dr Joan Steitz for suggesting circular templates for these studies, Joanne Meddick for technical assistance, Dr Aaron Shatkin for comments on the manuscript and Dr W. Szer for providing *E. coli*. This work was supported by NIH grant AI 15242.

MARILYN KOZAK

Department of Cell Biology,
New York University School of Medicine,
New York, New York 10016

Received 2 March; accepted 8 May 1979.

1. Kozak, M. *Cell* **15**, 1109-1123 (1978).
2. Kozak, M. & Shatkin, A. J. *J. biol. Chem.* **253**, 6568-6577 (1978).
3. Shatkin, A. J. *Cell* **9**, 645-653 (1976).
4. Kozak, M. & Shatkin, A. J. *Cell* **13**, 201-212 (1978).
5. Leung, D. W., Browning, K. S., Heckman, J. E., RajBhandary, U. L. & Clark, J. M. Jr *Biochemistry* **18**, 1361-1366 (1979).
6. Bretscher, M. S. *Nature* **220**, 1088-1091 (1968).
7. Bretscher, M. S. *J. molec. Biol.* **42**, 595-598 (1969).
8. Silber, R., Malathi, V. G. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3009-3013 (1972).
9. Singer, M. F. & Guss, J. K. *J. biol. Chem.* **237**, 182-189 (1962).
10. Kozak, M. & Shatkin, A. J. *J. biol. Chem.* **251**, 4259-4266 (1976).
11. Thach, R. E. & Doty, P. *Science* **147**, 1310-1311 (1965).
12. Bruce, A. G. & Uhlenbeck, O. C. *Nucleic Acids Res.* **5**, 3665-3677 (1978).
13. Kozak, M. & Shatkin, A. J. *Meth. Enzymol.* **60**, 360-375 (1979).
14. Subramanian, A. R. *Biochemistry* **11**, 2710-2714 (1972).

Evidence that *Euglena* chloroplasts do not export tRNAs

THE chloroplastic (Chl) and mitochondrial genomes from various organisms contain 20–30 tRNA cistrons. Several reports have demonstrated the presence of 20–30 tRNA species in these organelles. Thus, both chloroplasts and mitochondria seem to contain at least one tRNA species for each amino acid, plus initiator tRNA^{Met} and a few isoacceptors^{1–7}. In the few cases so far examined^{1,8}, organelle tRNAs have been shown to have primary sequences differing from those of their cytoplasmic isoacceptors. Three laboratories have demonstrated that *Euglena* Chl DNA codes for 22–26 species of tRNA^{9–11}. In the report of McCrea and Hershberger¹¹ total cellular tRNA was separated into two fractions (I and II) by dibutyl DEAE (DBAE) cellulose chromatography. Fraction I hybridised with both nuclear and Chl (~18 cistrons) DNAs. Fraction II seemed to hybridise only with Chl DNA (7–9 cistrons). Fraction II tRNA was not detected in isolated chloroplasts¹¹, but was found in either total cellular or cytoplasmic polysomal RNA¹², and contained acceptor activity for eight amino acids (Ala, Arg, Glu, Gly, Leu, Lys, Met and Phe)¹². As fraction II tRNAs hybridised with Chl DNA but were not found in isolated chloroplasts it was proposed that although they were transcripts of the Chl genome they were selectively exported to the cytoplasm^{11,12}.

In our laboratory chloroplast-localised tRNAs of *Euglena* have been identified by chromatographic comparisons with either total cellular tRNA or tRNA isolated from mutants lacking Chl DNA¹. For each of 16 amino acids studied, isolated chloroplasts were found to contain unique chromatographic species of tRNA¹. tRNA from isolated chloroplasts is complementary to 26 tRNA cistrons and Chl tRNA^{Phe} and tRNA^{Asp} are coded for by one cistron each⁹. In this report, we show that tRNAs obtained from isolated chloroplasts or whole cells hybridise to Chl DNA to the same extent and competition experiments reveal that these two sources of Chl tRNA complement identical cistrons. Furthermore, chromatographic evidence is presented that all the tRNA species represented as being totally exported to the cytoplasm¹² exist in isolated chloroplast preparations (which show no cytoplasmic contamination). These data indicate that there is no detectable transport of Chl tRNAs to the cytoplasm.

The tRNAs from whole cells and isolated chloroplasts were iodinated *in vitro*⁹. The amount of DNA complementary to each tRNA at infinite concentration was determined using double reciprocal plots (Fig. 1). Total cellular tRNA and tRNA from

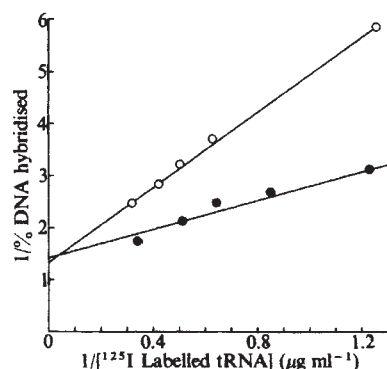


Fig. 1 Double reciprocal plot of the hybridisation of Chl tRNA (○) and total cellular tRNA (●) to Chl DNA. Filters containing 0.25 μg of Chl DNA were hybridised with Chl ¹²⁵I-labelled tRNA (9.7×10^6 c.p.m. per μg) or total cellular ¹²⁵I-labelled tRNA (1.4×10^7 c.p.m. per μg) in the presence of a fourfold excess of Chl rRNA as described previously⁹. At infinite concentrations of tRNA, 0.73% and 0.69% of Chl DNA is complemented by total cellular and Chl tRNAs, respectively.

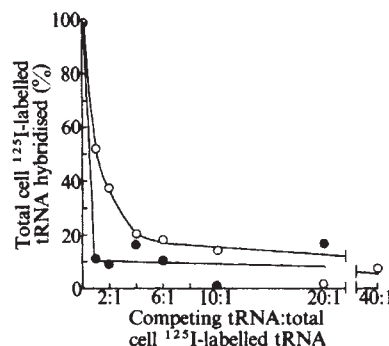


Fig. 2 Competition between Chl tRNA and total cellular tRNA for hybridisation with Chl DNA. Saturating concentrations of total cellular ¹²⁵I-labelled tRNA (9.4×10^7 c.p.m. per μg) were hybridised with filters containing 0.25 μg Chl DNA and increasing amounts of unlabelled total cellular (○) or Chl (●) tRNAs. All vials contained unlabelled Chl rRNA.

isolated chloroplasts are complementary to essentially the same percentages of Chl DNA (0.73% and 0.69, respectively). These data agree with our previous report⁹ that 0.74% of the Chl genomes hybridises with tRNA from isolated chloroplasts. Thus, tRNA from either source is complementary to 24–26 cistrons. Competition experiments were also carried out to verify that both total cellular tRNA and tRNA from isolated chloroplasts hybridised with the same portions of the Chl genome. A constant amount of total cellular ¹²⁵I-labelled tRNA was hybridised with Chl DNA in the presence of increasing amounts of unlabelled tRNA from either whole cells or isolated chloroplasts (Fig. 2). Both sources of tRNA are equally effective as competitors; over 90% of the hybridised counts could be competed for by the addition of unlabelled tRNAs. On a weight basis, tRNA extracted from isolated chloroplasts is a more effective competitor than tRNA from whole cells, because only ~35% of the total cellular tRNA is chloroplastic⁹. Thus, both whole cells and chloroplasts seem to contain an identical complement of Chl DNA-coded tRNAs.

Although McCrea and Hershberger reported¹² that fraction II tRNA contains acceptor activity for eight amino acids, previous reports have demonstrated that chloroplasts contain a unique species of tRNA^{Phe} (refs 13–16) and Kislev *et al.*¹⁷ demonstrated that *Euglena* chloroplasts contain acceptor activities for at least 16 amino acids including all those present in fraction II tRNA. Furthermore, we have been able to demonstrate (using BD-cellulose chromatography) that for each amino acid acceptor activity reported to be present in fraction II tRNA¹² a chromatographically unique species is present in isolated chloroplasts. Examples of this are shown in Fig. 3 for tRNA^{Ala} and tRNA^{Arg}. Isolated chloroplasts contain only a single symmetrical peak of alanyl and arginyl acceptor activity, whereas tRNA from whole cells has an additional isoacceptor for each of these amino acids. In whole cell preparations, Chl tRNA^{Ala} elutes as a trailing shoulder of the cytoplasmic isoacceptor and Chl tRNA^{Arg} also elutes later than its cytoplasmic counterpart. There is no evidence for the existence of any tRNA species in any case other than those from isolated chloroplasts or from the cytoplasm (nuclear coded).

The results presented here and those in many other reports are inconsistent with the hypothesis that the chloroplasts of *Euglena* selectively export a specific class of Chl-coded tRNAs that may be found in the cytoplasm or cytoplasmic polysomes^{11,12}. In the absence of chromatographic characterisation of the tRNA isoacceptors found in their DBAE-cellulose, cytoplasmic polysomal, total cellular or Chl fractions it is difficult to explain the results of McCrea and Hershberger.

Mutants lacking Chl DNA, and thus unable to code for Chl tRNAs, exhibit normal growth and are still able to synthesise all chloroplast-localised (nuclear-coded) enzymes so far examined¹⁸. Thus, there seems to be no physiological need for

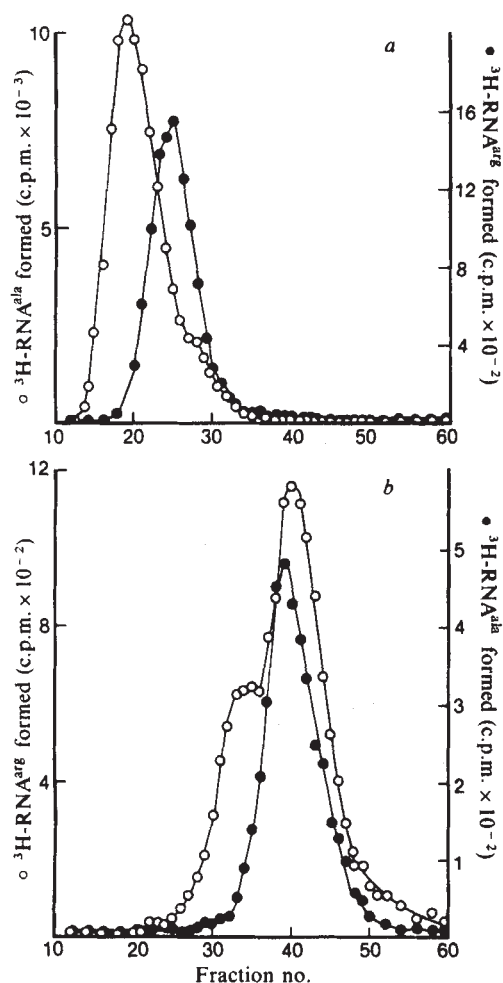


Fig. 3 Benzoylated-DEAE cellulose chromatography of total cellular (O) and Chl (●) deacylated tRNAs. 200 A_{260} units of tRNA from purified chloroplasts or 300 A_{260} units of tRNA from whole cells were applied to a 1×21 cm column and eluted with an 800-ml linear gradient from 0.33–1.25 M NaCl (ref. 16). Ala (a) and Arg (b) acceptor activities were assayed as previously described for Phe^{16,18} with the exception that the buffer used was HEPES, pH 7.5.

exported tRNAs as either constitutive or regulatory components of the cytoplasmic translation system. Therefore, we conclude that all evidence for the export of Chl DNA-coded tRNAs from chloroplasts is of doubtful significance.

STEVEN D. SCHWARTZBACH

School of Life Sciences,
University of Nebraska,
Lincoln, Nebraska 68588

W. EDGAR BARNETT

Biology Division,
Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37830

LANNY I. HECKER

Chemical Carcinogenesis Program,
Frederick Cancer Research Center,
Frederick, Maryland 21701

Received 2 April; accepted 15 May 1979.

1. Barnett, W. E., Schwartzbach, S. D. & Hecker, L. I. *Prog. nucleic Acid Res. molec. Biol.* **21**, 143–179 (1978).
2. Martin, R. P., Schneller, J. M., Stahl, A. J. C. & Dirheimer, G. *Nucleic Acids Res.* **4**, 3497–3510 (1977).
3. Kroon, A. M. *et al. Biochim. biophys. Acta* **478**, 128–145 (1977).
4. Van Ommen, G. J. B., Groot, G. S. P. & Borst, P. *Molec. gen. Genet.* **154**, 255–262 (1977).
5. Martin, N. C. & Rabinowitz, M. *Biochemistry* **17**, 1628–1634 (1978).
6. De Vries, H. *et al. Biochim. biophys. Acta* **520**, 419–427 (1978).
7. Malnoé, P. & Rochemaix, J. D. *Molec. gen. Genet.* **166**, 269–275 (1978).
8. Martin, R. P. *et al. Nucleic Acids Res.* **5**, 4579–4592 (1978).

9. Schwartzbach, S. D., Hecker, L. I. & Barnett, W. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1984–1988 (1976).
10. Groul, D. J. & Haselkorn, R. *Biochim. biophys. Acta* **447**, 82–95 (1976).
11. McCrea, J. M. & Hershberger, C. L. *Nucleic Acids Res.* **3**, 2005–2018 (1976).
12. McCrea, J. M. & Hershberger, C. L. *Nature* **274**, 717–719 (1978).
13. Barnett, W. E., Pennington, C. J. & Fairfield, S. A. *Proc. natn. Acad. Sci. U.S.A.* **63**, 1261–1268 (1969).
14. Reger, B. J., Fairfield, S. A., Epler, J. J. & Barnett, W. E. *Proc. natn. Acad. Sci. U.S.A.* **67**, 1207–1213 (1970).
15. Fairfield, S. A. & Barnett, W. E. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2972–2976 (1971).
16. Hecker, L. I., Uziel, M. & Barnett, W. E. *Nucleic Acids Res.* **3**, 371–380 (1976).
17. Kislev, N., Selsky, M. I., Norton, C. & Eisenstadt, J. M. *Biochim. biophys. Acta* **287**, 256–259 (1972).
18. Bovarnick, J. G., Schiff, J. A., Freedman, Z. & Egan, J. M. *J. gen. Microbiol.* **82**, 63–71 (1974).

Molecular basis for proton-dependent anion binding by soybean leghaemoglobin *a*

HAEM proteins are widespread in nature and exhibit diverse biological activities. The properties and function of a given haem protein depend critically on the environment of the haem group. The reactivity of the haem is often influenced by one or more haem-linked ionisable groups. Thus, oxygen affinities of mammalian haemoglobins, for example, exhibit considerable pH dependence (the Bohr effect¹). Anion binding to some haem proteins is also a proton-dependent reaction (for example, peroxidase², catalase² and leghaemoglobin^{3,4}) whereas for others (myoglobin, haemoglobin¹) it is essentially independent of pH. Identification of the ionisable groups which influence the ligand-binding reactions of haem proteins is of fundamental importance in the elucidation of structure/function relationships. Nuclear magnetic resonance (NMR) spectroscopy is one of the most direct methods for observation of titrating groups in proteins. We report here on NMR studies of the nicotinate complex of soybean ferric leghaemoglobin *a*. On the basis of these studies we propose a role for a haem propionic acid group as a mediator of anion binding.

All experiments were carried out with soybean ferric leghaemoglobin *a* prepared by a previously described procedure⁵. Reactions with acetate to form a high-spin complex³ and nicotinate to form a low-spin ferric complex⁴ are proton dependent, with pK values of 4.8 and 4.9, respectively. The affinity of leghaemoglobin for these anions increases with decreasing pH, suggesting the presence of a haem-linked ionisable group which, when protonated, facilitates access or binding of anions to the haem.

The low field region of the proton NMR spectrum of ferric leghaemoglobin nicotinate is shown in Fig. 1a. Resonances in this region of the spectrum are subject to large hyperfine shifts due to the presence of the single unpaired electron of the iron atom and generally arise from protons on the porphyrin ring or axial iron ligands⁶. The resonances are relatively narrow due to the rapid electron spin relaxation characteristic of low-spin ferric haem⁶. The resonances at 25.6 and 18.5 p.p.m. at 45 °C and pH 4.5 can be unambiguously assigned to two of the four porphyrin ring methyl groups, on the basis of their intensities and previous NMR studies of haem proteins and model haem complexes⁶. The two single proton resonances at 14.9 and 12.7 p.p.m. at pH 4.5 are assigned to the two α -CH₂ protons of one of the haem propionic acid side chains. The resonances shift at acid pH, titrating with a pK of 4.9 (Fig. 2). The assignment is confirmed by the observation of a large negative nuclear Overhauser effect^{7,8} (~40%) on each resonance when the other is irradiated (Fig. 1b, c). The protons are therefore in close spatial proximity such that each dominates the nuclear relaxation of the other, as would be expected for two inequivalent protons of a propionic acid methylene group. Restriction of the rotational motions of the propionate side chain would give rise to magnetic inequivalence of its α -CH₂ or β -CH₂ protons. Assignment is made to the α -CH₂ rather than the β -CH₂ protons on the basis of the larger contact shift experienced by the former. No other protons on the haem or axial iron ligands can give rise to these

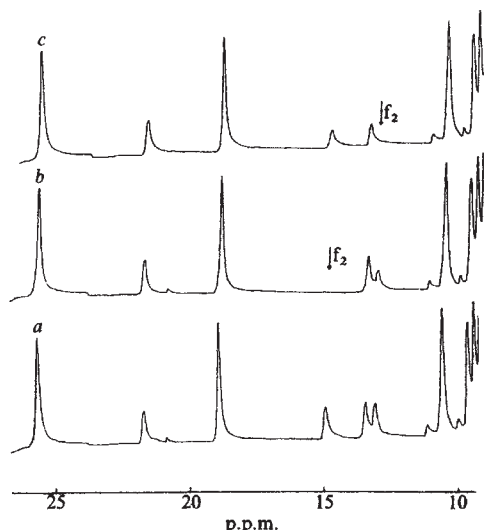


Fig. 1 *a*, Low field region of the 270-MHz proton NMR spectrum of soybean ferric leghaemoglobin *a* nicotinate at 45 °C and pH 4.5. *b*, Negative nuclear Overhauser effect (~40%) on resonance at 12.7 p.p.m. due to irradiation at 14.9 p.p.m. *c*, Negative nuclear Overhauser effect (~40%) on resonance at 14.9 p.p.m. due to irradiation at 12.7 p.p.m. Spectra were recorded using a Bruker HX-270 spectrometer operating in the pulsed Fourier transform mode. The residual HOD peak was suppressed by selective gated irradiation. Overhauser effects were measured by selective irradiation which was gated off immediately before data acquisition. Dioxan was used as an internal standard but all peaks are referred to trimethylsilylpropanesulphonic acid. Solutions for NMR were 3–4 mM in leghaemoglobin in phosphate buffer (10 mM, in D₂O). The pH was adjusted by careful addition of DCl or NaOD. The pH values quoted are uncorrected meter readings.

resonances. We believe this to be the first unequivocal assignment of the α -CH₂ proton resonances of a haem propionic acid side chain in the NMR spectrum of a haem protein. The assignments of the other resonances in the low field region will be published later.

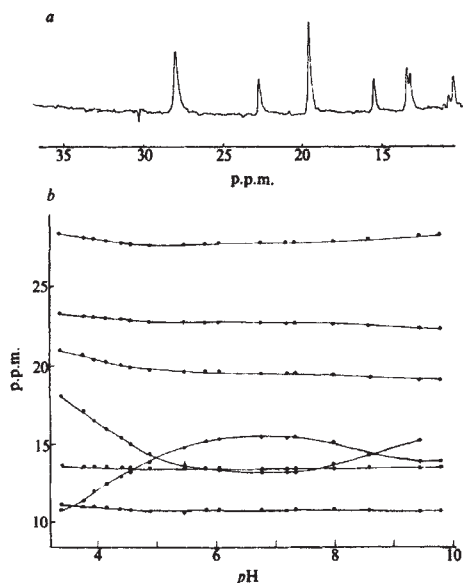


Fig. 2 *a*, Proton NMR spectrum of soybean ferric leghaemoglobin *a* nicotinate at 25 °C and pH 7.2. *b*, pH dependence of the low field region of the NMR spectrum. Exchange between nicotinate and ferric leghaemoglobin is slow on the NMR time scale. An excess of nicotinate was used such that the leghaemoglobin was fully complexed at pH values below ~7.0. At higher pH, dissociation of nicotinate leads to a decrease in intensity of the resonances.

Unfortunately, we are unable to assign the resonances of this haem propionic acid side chain in the spectrum of ferric leghaemoglobin acetate. These and the resonances of other protons near the haem are very broad due to the slow electron spin relaxation of the high-spin ferric haem.

The propionic acid resonances in the spectrum of leghaemoglobin nicotinate reflect the titration of two ionisable groups (Fig. 2). As the pH is increased above ~7.0 it is evident from the observed decrease in intensity of the low field haem resonances that nicotinate binding becomes weaker. The large perturbations at acid pH (Fig. 2) clearly arise from titration of the propionate side chain carboxyl group with *pK* 4.9. A further substantial perturbation is observed above pH 7. Its magnitude suggests a close interaction between the haem propionate and a group with *pK* 8.3. We identify this other group as the distal histidine residue, as we can follow the titration of a singlet resonance with *pK* 8.3 as it shifts from 8.1 p.p.m. to ~7.2 p.p.m. with increasing pH (unpublished observations). The high *pK* of this histidine is consistent with formation of a salt link or hydrogen bond to the carboxyl group of the nicotinate ligand^{9,11}.

On the basis of affinity labelling with mesohaem sulphuric anhydride, Ellfolk and Perttola¹⁰ have suggested that one of the haem propionic acid side chains forms a salt bridge to Lys 57 in soybean ferric leghaemoglobin *a* at low temperature. Our NMR studies provide no evidence to support this proposal but indicate an interaction between one of the haem propionates and the distal histidine in leghaemoglobin nicotinate.

The present NMR studies indicate that at least one of the haem propionic acid substituents in soybean ferric leghaemoglobin *a* nicotinate is accessible to the solvent and free to titrate with a *pK* of 4.9. Reactions of the ferric protein with acetate and nicotinate are proton dependent, with *pK* values of 4.8 and 4.9, respectively^{3,4}. These *pK*s correspond closely to that of the free propionic acid side chain, which is thus likely to be the group which mediates anion binding. We thus propose a model for the reactions of soybean leghaemoglobin *a* with anions in which a haem propionic acid group, when deprotonated, functions as an electrostatic 'gate' which restricts the access of anions to the haem pocket. Protonation of this group would facilitate anion binding. In an alternative mechanism access of anions to the haem is sterically hindered as a result of protein conformational changes accompanying deprotonation of the haem propionic acid. This mechanism may be discounted, as reactions of leghaemoglobin with neutral ligands such as pyridine⁴ and nicotinamide (C.A.A., unpublished observations) are independent of pH.

We thank Mrs J. Wellington for assistance and the Australian Research Grants Committee for financial support and for the use of facilities at the National NMR Centre.

JILL TREWHELLA
PETER E. WRIGHT

Department of Inorganic Chemistry,
University of Sydney,
Sydney, NSW 2006, Australia

CYRIL A. APPLEBY

Division of Plant Industry,
CSIRO,
Canberra, ACT 2601, Australia

Received 9 April; accepted 8 May 1979.

1. Antonini, E. & Brunori, M. *Hemoglobin and Myoglobin in their Reactions with Ligands* (North-Holland, Amsterdam, 1971).
2. Pratt, J. M. in *Techniques and Topics in Bioinorganic Chemistry* (ed. McAuliffe, C. A.) 106–204 (Macmillan, London, 1975).
3. Ellfolk, N. *Acta chem. scand.* **15**, 975–984 (1961).
4. Appleby, C. A., Wittenberg, B. A. & Wittenberg, J. B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 564–568 (1973).
5. Appleby, C. A., Nicola, N. A., Hurrell, J. G. R. & Leach, S. J. *Biochemistry* **14**, 4444–4450 (1975).
6. Wuthrich, K. *Struct. Bonding* **53**, 53–121 (1970).
7. Noggle, J. H. & Schirmer, R. E. *The Nuclear Overhauser Effect* (Academic, New York, 1971).
8. Balaram, P., Bothner-By, A. A. & Dadok, J. *J. Am. chem. Soc.* **94**, 4015–4017 (1972).
9. Johnson, R. N., Bradbury, J. H. & Appleby, C. A. *J. biol. Chem.* **253**, 2148–2154 (1978).
10. Ellfolk, N. & Perttola, U. *Biochim. biophys. Acta* **534**, 153–159 (1978).
11. Harutyunyan, E., Kuranova, I., Grebenko, A., Voranova, A. & Vainshtein, B. *Proc. XIth IUPAC Int. Symp. Chem. Natural Products* (eds Marekov, N. et al.) **2**, 477–481 (1978).

reviews

Balanced overview of cancer research

Richard Peto

Cancer, Science and Society. By John Cairns. Pp. 199. (Freeman: San Francisco and Reading, 1979.) Hardback £6.20; paperback £2.90.

THE general public, which provides the money that cancer research consumes, seems to hope for the eventual emergence of the sort of "magic bullet" cancer cure which is as easy, specific and effective as an antibacterial agent can be against infections. In reality, there is little hope of any such thing; indeed, the prospect is so remote that it scarcely enters into the day-to-day conversation of most cancer research workers. Researchers talk of slightly earlier detection, of moderate improvements in current methods of therapy, of the prevention of minor environmental carcinogens, of the identification of life styles (with less meat, less tobacco and lots of vitamins) which would, if only anyone would adopt them, somewhat reduce cancer incidence rates. They talk of developing a greater understanding of tumour antigens or of the nature of the transformed state, but easy and specific cures are not sufficiently realistic to be of interest.

In this book, John Cairns tries to explain the realities of cancer biology and epidemiology in sufficient detail to be interesting and to let the facts themselves give the reader a realistic perspective on "cancer, science and society". In this I feel he succeeds excellently, and anybody with a background in any area of science would enjoy this overview. In style it reminds me of the *Scientific American*; it is clearly written and well illustrated, with separate chapters on each of several completely different areas of research, from demography to biology.

What is perhaps unusual in a book by a professional cancer research worker is his whole-hearted public acceptance of the idea that we cannot expect any "magic bullet" cancer cure to be discovered for at least several decades. Whether this pessimism is justified or not, it is certainly a view which deserves to be more seriously debated than in the past. Obviously, an optimistic claim that understanding viruses, interferon, immunology, and so on, will soon offer specific cures is impossible to refute and, if accompanied

by any new finding, will be more newsworthy and therefore more widely published than the pessimistic claim that we are decades away from a detailed enough understanding of cell biology to devise a cancer cure rationally. However, this inevitable emphasis on optimistic perspectives may in the long run produce the sort of distorted hopes of rapid victory that, as with wartime propaganda, is fostered by reporting victories more prominently than defeats. If Cairns is right in his pessimism about the next few decades of cancer therapy, then he is surely right in his inference that *clinical* cancer research nowadays is over-supported, and that the two chief areas of research to foster over the next few decades should be cancer prevention and, at the other extreme, fundamental biological research with no requirement for any obvious connection with cancer. As a strategy, it may well be more successful than seeking vast amounts of money for clinical research with the slogan "Conquer cancer in the seventies".

Whether or not this overall perspective is accepted, the book provides an interesting, readable and balanced overview of cancer research, written for the sort of intelligent general public who perhaps do not know very much about cancer research but who might feel sufficiently interested to read a whole issue of the *Scientific American* devoted to the subject. As an introduction to the subject, it is as good as any I have seen, and would be an excellent textbook for an undergraduate or post-graduate biology course on cancer. However, the people who I chiefly hope will spend a pleasant evening reading it are the cancer research workers who are concentrating on particular details and who need some feel for how (or whether) their activities make sense in a wider context.

The book begins with demography. As a cause of death, cancer was much less important in the nineteenth century, simply because cancer tends to be a disease of old age, and most people never reached old age. However, during the past 100 years life expectancy in developed countries has suddenly doubled, and over 75% of people will now live beyond the age of 65.

Consequently, the present chief cause of death are those which act in old age, the most prominent being vascular disease (half of all deaths), cancer (one-sixth of all

deaths) and, in Britain but not America, respiratory disease (another one-sixth). On average, those who die of cancer would only have lived another 12 years if their cancer had been prevented or cured, which is, of course, likely to be a worthwhile gain for the individual if a fair number of those years were passed in reasonable health. However, an average of 12 years extra for one-sixth of the population only represents an additional two years of life expectancy for the population as a whole, and is not comparable with the 30 or more years of extra life expectancy that has been conferred by control of the diseases which used to kill so many children and young adults. (Only 1% of cancers affect children.) In short, even the total prevention or cure of all cancer cannot yield the sort of huge change in the human condition that has followed from adequate nutrition and from the control of the major infectious diseases.

Turning from demography to epidemiology, cancers of different anatomical sites (for example, stomach and lung) behave as different, and largely unrelated, human diseases. Causes of one are unlikely to be causes of the other; a genetic predisposition toward one does not imply an abnormal risk of the other (except in a few people with very unusual genetic defects); and, most importantly, control of one type of cancer does not mean that the age-specific risk from any other cancer need be increased. Thus, there is not one cancer problem but many.

The converse of this rather unwelcome conclusion is that although the total age-specific cancer death rates in different countries tend to be rather similar, this overall similarity is the sum of many widely different parts. Indeed, one of the most encouraging findings in the whole of cancer research is that there exist very great differences between countries in the age-specific risks from particular cancers. (The risk of developing cancer increases so sharply with age that it is often misleading to talk about overall probabilities of cancer, for in poor countries where most people still do not reach old age the net likelihood of cancer is small even if the age-specific cancer risks are normal.) For example, lung cancer is 35 times as common among people of a given age in Britain as among people of that age in

Nigeria, suggesting that the large majority of our 35,000 annual lung cancers are preventable. Stomach cancer is 25 times commoner in Japan than in Uganda but has been decreasing quite rapidly over recent decades as Japan has adopted a richer, more Western diet. Cancer of the oesophagus is more than 100 times commoner in north-eastern Iran (where it is the major cause of death) than in most other parts of the world, and has apparently been common there for centuries. The general rule is that, although some cancers are rare everywhere, every cancer which is common in one country is rare in another country, and so for every common cancer there presumably exist life-styles which humans would find acceptable that would make it vastly less common, probably without making any other type of cancer much more common. This presumption is supported by what sparse data are available on rapid changes in the age-specific incidence rates of various cancers over the past few decades. In America in 1930, cancer of the stomach was the commonest fatal cancer, whereas cancer of the lung was rare. Nowadays, the opposite is true. Over the same period in Britain, cancer of the tongue has decreased twentyfold, whereas cancer of the lung has increased twentyfold.

Thus, cancers have causes. But what might they be? It is tempting to ascribe them to large-scale chemical pollution of air, food or water, but this is probably mistaken, if only because the widespread dissemination of chemicals (in plastics and agriculture) is too recent for its effects on cancer yet to be apparent. In Britain 30%, and in America 25%, of all cancer deaths can be attributed to smoking, and if those cancers which are closely associated with smoking are omitted there is no upward trend in the overall death rates from the remaining cancers. If smoking accounts for only just over a quarter of all cancer deaths, what preventable causes account for the rest? An unproven but plausible hypothesis ascribes a large role in the genesis of colorectal, breast and endometrial cancer to meat, or fat, consumption; and Cairns presents the visually striking relationship between meat consumption and colon cancer. (Unfortunately, he does not also discuss the interesting hypothesis that certain vegetables or vitamins may be actively protective.)

Cairns then turns his attention to the world of cancer research. It is a rather aristocratic overview ("Biology and cancer research have developed together. Invariably, at each stage the characteristics of the cancer cell have been ascribed to some defect in whatever branch of biology happens at the time to be fashionable and exciting; today it is molecular genetics . . ."), but tends to be the more interesting for being so.

In order to understand the cancer cell we must be taught something of the chemical

basis of heredity, and this is done nicely from first principles, starting with DNA, RNA, the triplet code and the effects of mutations in it. Repair mechanisms are described, and then Cairns is ready to move on to the experimental production of cancers by chemical carcinogens, radiation and viruses. Many of the chemical agents which produce cancers in rodents also turn out to be mutagens for bacteria (or chemicals which can be converted into bacterial mutagens by enzymes which are present in normal mammalian cells). Obviously, this offers the possibility of detecting certain carcinogens quickly by allowing them to act on a mixture of mammalian cell components and bacteria,



Oliver escapes being bound apprentice to the Sweep

Cancer of the scrotum was prevalent in boys not so lucky

and seeing whether they cause mutations in the bacteria. Such methods may, however, be of more use in the prevention of new carcinogens than in detecting those which have already affected us for decades.

The two chapters on the experimental production of cancers are among the best in the book, perhaps because Cairns is here reviewing his own field. Perhaps because I am not, or perhaps because they already represent a greatly condensed account of a great deal of material, I find them difficult to summarise.

The concept of the metabolic activation of relatively inert precursors into electrophilic species that can bind to DNA took a long time to evolve, but it has allowed a grand unification to emerge of the mode of carcinogenic action of a wide class of 'initiating' agents, including chemicals that do and chemicals that do not need metabolic activation, some aspects of radiation carcinogenesis, and even certain viruses which seem simply to cause non-specific damage to DNA. The notable exceptions to this grand unification are the "promoting" agents, which seem to bind only to the surface of the target cell but which can greatly increase the carcinogenic effects of 'initiators', and the specific tumour viruses, which are described nicely in detail.

No virus in its right mind wants to exterminate its host; viruses just want to

make sure that host animals continue to excrete enough virus particles throughout their lifespan to make them infect all their acquaintances. The best way to do this is for the virus to insert its genetic material into the DNA of the host's own germ cells or, failing that, the host's tissues, in such a way that copies can be run off whenever necessary. In the process, the virus may by mistake trigger certain host cell functions which make the host cell behave like a cancer cell, but from the virus's point of view that is an expensive mistake, as it kills the host. Alternatively, a really unfortunate virus may when it releases its own genetic material from the host's material inadvertently include a little bit of host DNA containing an instruction which, if expressed without proper host control, could cause uncontrolled proliferation. Without the advent of cancer research workers (who are pleased to discover and to selectively propagate such oddities), such viruses would be at an evolutionary dead end, and their non-sarcomagenic cousins, who could live in hosts without killing them, would meekly inherit the earth.

In short, Cairns is not sure which, if any, of the known tumour viruses are laboratory artefacts, perhaps useful as tools for controlled manipulation and studying the 'transformed state' *in vitro* but not as models of human carcinogenesis.

Nor is he sure, of course, what the essential attributes of the "transformed state" are; it seems as though the key tricks needed to grow the cells one really wants to study have still not been found. If, as is often done, we define 'transformation' as the ability to grow indefinitely in soft agar, 'transformed' cells will indeed usually develop into a tumour when transplanted into a host. However, cells from real human tumours will usually not grow at all in soft agar or in any other *in vitro* conditions yet tested, and moreover normal human cells growing *in vitro* are almost impossible to 'transform', so most studies of transformation are done with rodent cells. These 'transform' spontaneously, or in response to viruses, chemicals or radiation, but although the biochemistry of transformation is currently being elucidated, its relevance is not yet clear.

Returning to matters which are better understood, the many decades that may elapse between the action of an initiator and the eventual emergence of a carcinoma are discussed in terms of multistage models, emphasising the relevance of such ideas for early detection programmes. The final chapter on prospects for the future is almost too pessimistic about whether preventive measures, even if discovered, will be implemented, and emphasises very strongly the improbability of discovering any simple, effective and specific cancer cure in the foreseeable future. □

Richard Peto is Reader in Cancer Studies at the University of Oxford, UK.

Discrete charges in quantum mechanics

Gravitation, Electromagnetism and Quantized Charge: The Einstein Insight. By Daniel M. Pisello. Pp. 94. (Ann Arbor Science: Ann Arbor, Michigan, 1979.) Distributed in UK, Africa and Europe by Wiley: Chichester, UK. £7.55.

By means of dogmatic pronouncements on the limitations of human imagination and language, the quantum mechanical philosophy effectively takes away from the younger physicists their right to question and create. If one finds it necessary to question or theorize beyond the restraints so successfully built into the quantum mechanical rationale, one is then, by fiat, pejoratively considered a revolutionary". Thus ends Mr Pisello's thin treatise, half philosophical, half mathematical, and always bitter. This work will not revolutionise physics, "pejoratively" or otherwise. The book's interest, however, lies in the way its author attempts to link his philosophical and political attitudes with a mathematical search for a purely field theoretical approach to the existence of discrete charges in quantum mechanics. The result is an exercise in topology elevated to the rank of "unified theory" through a casual parallel with Einstein's own epistemological objections to the theory of quantum mechanics.

The first three chapters present the conceptual difficulties arising from scientific breakthroughs like the theories of special relativity, general relativity and quantum mechanics. Old and new concepts which apparently contradict each other are reviewed, the limitations inherent in the use of a language created within an earlier framework are discussed, and the need for a "principle of complementarity" is challenged on philosophical grounds. A rapid analysis of the double slit experiment for which "neither linear wave nor discrete particle interpretation provides an adequate representation of the transfer of energy from source to detector" is used to justify searching for the "basic nature of the field variable". The remainder of the book is devoted to this search in the form of mathematical exercises in homotopy theory and general relativity. It ends with a succinct theory of charge as a counting procedure for "an identifiable number of separate left- and right-hand doughnuts".

Noticing that physical field theories deal only with continuous entities while the physical world displays basic discontinuities (the elementary electric charge of the electron being the one specific case considered here) is not new, but the whole preoccupation with continuity, infinities and singularities is indeed continuing. In this sense Mr Pisello's book is part of the contemporary mood. To try

to save continuity by hiding the "inside" of particles into topology has been done with mastery by Professor John Wheeler creating "wormholes" for the delight of many students of general relativity. Mr Pisello attempts something similar but each piece of his puzzle unfortunately remains physically and also mathematically quite disconnected. At the Einstein centennial celebration at UNESCO in Paris a few weeks ago, Professor Dirac stressed his own dis-satisfaction with quantum mechanical calculations for which we "learned to turn a disinterested eye to infinities" and urged that Einstein's objections to quantum mechanics be seriously considered anew. Mr Pisello does just that in so far as he retakes "Einstein's insight" in the direction of unified field theories. He fails, however, to display a satisfactory theory, thus joining many others in this particular domain. The author also clearly shares with Einstein a socialist point of view upon the world; this does not justify the bloated language used

by the publishers in the presentation of the book: "Pisello agrees with Einstein that quantum mechanics is a reactionary theory devised to maintain the mechanical view". Trying to get too close to Einstein, you may burn yourself up.

Failing to present clearly physical ideas and political analysis may be a consequence of the deep social frustrations experienced by Mr Pisello's generation of physicists. Trained in high energy particle physics, they got their degrees in America at the time of the Vietnam war. Their belief in 'pure' science was shaken when many of their peers were exposed as direct advisors for the US Department of Defence, and they could not get steady work as physicists in a tight competitive job market. This generation of physicists remains anxiously in search of the social and historical connections of scientific work.

Maurice Bazin

Maurice Bazin is Visiting Professor of Physics at the University of Campinas in Brazil.

Sociology of science

Science and the Sociology of Knowledge. Controversies of Sociology. Vol 8. By M. Mulkay. Pp. 132. (George Allen & Unwin: London, 1979.) £7.50; paperback £3.50.

FROM the time of the 'founding fathers' of sociology in the nineteenth century, sociologists have been interested in the way in which bodies of knowledge are influenced by the social and cultural contexts in which they are produced. This is the sociology of knowledge. But scientific knowledge was always seen as something apart, special, and essentially independent of external social influences in a way that was not true of other specialised bodies of thought and knowledge. Thus, when a sociology of science emerged, and developed (quite rapidly from the 1950s), it became something other than the attempt to explain the content of scientific knowledge in terms of social structures and processes. The sociology of science developed as the attempt to understand the social conditions which made scientific progress possible: that is, it was concerned with general social and political conditions (giving rise to the idea of correspondence between science and democracy), as well as with the way in which scientific communities worked. There was a lot of interest in the 'normative structure' of science: the rules or moral code governing the ways in which scientists worked and interacted with each other.

A good deal of work in sociology of science is still being done along these lines, some of the most interesting of it

concerned with the emergence and functioning of scientific communities in smaller or less industrialised nations. But Michael Mulkay is concerned with something else. His argument is that both this approach as well as discounting of the possibility of a sociology of scientific knowledge are founded upon a particular philosophical view of the nature of science, which he calls "the standard view of scientific knowledge". Roughly speaking, this posits a real, objective, and uniform natural world; it assumes that science, through unbiased and detached observation, can formulate laws of nature which describe this world; that science makes use of entirely objective criteria for evaluating knowledge claims, such that science is ultimately free of bias; and that theoretical laws are of a different status to empirical ones and stand in a quite different relationship to the natural world. He goes on to argue that this philosophy of science is no longer acceptable, and that it is therefore legitimate to try to establish a sociology of scientific knowledge.

Mulkay characterises the "new philosophy of science" as showing the dependence of all observation on theory, of all factual statements on (arbitrary) linguistic frameworks. All meaning in science becomes the product of negotiated agreement and is embedded in some changing framework of values and assumptions. Feyerabend's "epistemological anarchism" is an extreme version of all this.

There have now been a few sociological studies of science which made use of this relativised version of science, and Mulkay describes some of them. It becomes important to try to understand how scientists 'negotiate' 'what counts' as the

replication of an experiment (for example). Other work has explored, most interestingly, the way in which scientific disputes are conducted, and the way in which all kinds of political resources and non-scientific opinion can be used.

The final part of the book deals with the exchanges between science and other aspects of culture. This includes the use which may be made of theoretical perspectives or forms of explanation available in general thought. Unfortunately there is little discussion of the conditions under which these will be drawn on.

This whole undertaking is relatively new in sociology, and one cannot yet say what it

will yield. My own view is that there are more new philosophies of science than Mulkay implies and that there will gradually emerge many attempts at quite different sociologies of scientific knowledge. Certainly the work which has been done so far is interesting, and could give sceptical scientists pause for thought. The book is likely to be of interest mainly to those who have already read into the field of sociology of science.

Stuart S. Blume

Stuart S. Blume is Senior Research Fellow in the Department of Social Science and Administration, London School of Economics, University of London, UK.

Mysterious Solar System

Mysterious Universe: A Handbook of Astronomical Anomalies. By W.R. Corliss. Pp. 710. (The Sourcebook Project: P.O. Box 107G, Glen Arm, Maryland 21057, 1979.) \$15.95; £8.

MYSTERIES fascinate me and this book contains an abundance. It is a classical "What done it?" — the inanimate scientific relation to the "Who done it?" William R. Corliss has scoured a host of scientific journals for reports of astronomical oddities and has assembled over 500 short articles and extracts, on a multitude of subjects ranging from aurorae in comet tails to zodiacal light-rifts and black and white holes to X-ray bursts.

It is strong meat. Taken at one go the unceasing succession of anomalies, unknowns, mysteries and strange sightings leave the reader mentally punch drunk. You will stagger from the chair to the window to reassure yourself that the Sun and planets and bright stars are still where you thought they should be and have not all capriciously shuffled themselves. But in short bursts the book is a joy and the reader's mind is continually accelerated away in pursuit of ideas and possibilities that may account for the strange events being described.

The timespan covered by the book is reasonably broad, the most up-to-date extracts coming from mid-1978. Approaching from antiquity the number of articles verses time curve increases sharply around the middle of the 19th century, a time which seems to be governed more by the birth of journals such as *Nature* and *Scientific American* than by a sudden onset of mysterious happenings or human curiosity. Our excitement over the 'recent' discovery of the rings of Uranus pales somewhat when we read that Challis and Lassell reported an observation of a ring in 1847. These gentlemen also observed a ring around Neptune.

And things disappear in astronomy too.

Vulcan the supposed intra-Mercurial planet sprang into prominence in 1876 only to fizzle from view in the early 1900s. Corliss collects together 14 papers on this subject. We are also introduced to Neith the lost satellite of Venus. The eighteenth-century observations of this object remain enigmatic and simply accumulate distain as they become eroded by time. We can easily smile condescendingly and talk about 'ghosts' in ancient telescope eye pieces but maybe...

Simple classical vibrator

The Physics of Vibration. By A. B. Pippard. Vol. 1. (Cambridge University Press: Cambridge, London, New York and Melbourne, 1978.) £22.50.

THIS is the first volume, devoted to the simple classical vibrator, of a work on the physics of vibration. It is designed to bridge the gap between undergraduate textbooks on the subject and specialist treatises. It ranges widely, discussing not only the obvious examples from mechanics, elementary electronics and acoustics, but also nuclear magnetic resonance, cyclotron resonance and Josephson junctions. The subsequent volume or volumes will deal with specifically quantal phenomena and with vibrations of complex systems.

There is an introductory chapter describing very clearly the scope of this work; a chapter on the free vibrator; two chapters on applications of complex variables and on Fourier methods; three chapters on spectrum analysis, on the driven harmonic vibrator, and on waves and resonators; then one on velocity dependent forces groups, together with various examples such as whirling, the gyro-pendulum, nuclear magnetic resonance, cyclotron resonance and helicons. This chapter somewhat breaks the flow of the rest of the book, and does not fit in very happily; I am not convinced

Any collection of articles has the character of the collector strongly imprinted on it. This I'm sure cannot be avoided. Corliss shows a strong bias towards the Solar System. For example, the Moon, meteors and meteorites take up 26% of the book, the same amount of space as all the stars, the Sun, galaxies and cosmology put together. A similar percentage is taken up by the planets, leaving the zodiacal light, comets, planetary regularities and the environs of planet Earth to absorb the remainder.

I would be happier if the title of the book was *Mysterious Solar System*. Failing this, there are an enormous number of mysterious star types, stellar origin and evolution problems, interstellar, galactic and intergalactic matter and composition anomalies that could have been expanded to give the book a more 'universal' balance.

However, as a sourcebook of a lot of strange Solar System happenings and a number of stellar and galactic ones, the book is to be highly recommended. It has been well produced, extremely well indexed and is excellent value for money.

David W. Hughes

David W. Hughes is Lecturer in Physics and Astronomy at the University of Sheffield, UK.

that this level of treatment of magnetic resonance will be helpful either to the expert or to the inexpert. A chapter on the driven anharmonic vibrator discusses subharmonics and stability. There is a short chapter on parametric excitation and a chapter on maintained oscillators. The final chapter is on coupled vibrators, and includes the Huygens phenomenon, frequency-locking and superconducting weak links.

Although the majority of examples are taken from electric circuit theory and mechanical oscillators, there are many others, some quite intriguing. The level of discussion is what one would expect from a good experimental physicist. The main interest is in common patterns of behaviour rather than in the details of operation, yet the details are worked out for many special cases.

I think that is primarily a book for the teacher rather than for the student. It does not treat any one subject in sufficient detail to be a useful source for someone who really wants to understand it, but it is more detailed than most students who want a general impression of the subject would want. It is, however, full of valuable new insights and analogies which should be helpful to anyone who needs to teach this sort of material. No physics library should be without a copy.

D. J. Thouless

D. J. Thouless is Professor of Applied Science at Yale University, New Haven, Connecticut.